nonconform: Conformal Anomaly Detection (Python)

Oliver Hennhöfer [1]

1 Intelligent Systems Research Group (ISRG), Karlsruhe University of Applied Sciences (HKA), Karlsruhe, Germany

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

The Python package `nonconform` provides statistically principled uncertainty quantification for unsupervised anomaly detection. It implements methods from conformal anomaly detection (Bates et al., 2023; Jin & Candès, 2023; Laxhammar & Falkman, 2010) based on the principles of one-class classification (Petsche & Gluck, 1994). The ability to quantify uncertainty is a fundamental requirement for AI systems in safety-critical domains, where reliable decision-making is essential.

Based on the underlying principles of conformal inference (Lei & Wasserman, 2013; Papadopoulos et al., 2002; Vovk et al., 2005), `nonconform` converts raw anomaly scores from an underlying detection model into statistically valid p-values. This is achieved by calibrating the model on a hold-out set of normal data; the p-value for a new test instance is then calculated as the relative rank of its anomaly score compared to the scores from the calibration set. By framing anomaly detection as a series of statistical hypothesis tests, these p-values allow for the systematic control of the False Discovery Rate (FDR) (Bates et al., 2023; Benjamini & Hochberg, 1995) at a pre-defined significance level (e.g., $\alpha \leq 0.1$). The library integrates with the popular pyod library (Chen et al., 2024; Zhao et al., 2019), making it easy to apply these conformal techniques to a wide range of anomaly detection models.

Statement of Need

A primary challenge in anomaly detection is setting an appropriate anomaly threshold, which directly impacts the false positive rate. In high-stakes domains such as fraud detection, medical diagnostics, and industrial quality control, controlling the proportion of false positives is crucial, as frequent false alarms can lead to *alert fatigue* and render a system impractical. The `nonconform` package addresses this by replacing raw anomaly scores with p-values, which enables formal FDR control. This makes the conformal methods *threshold-free*, as decision thresholds are a direct result of respective statistical procedures.

$$FDR = \frac{\text{Efforts Wasted on False Alarms}}{\text{Total Efforts}}$$

(Benjamini et al., 2009; Benjamini & Hochberg, 1995)

Moreover, conformal methods are *non-parametric* and *model-agnostic*, making them compatible with any model that produces consistent anomaly scores. The `nonconform` package provides a range of strategies for creating the calibration set from training data, even in low-data regimes (Hennhofer & Preisach, 2024). With the gathered calibration set, the package can compute standard conformal p-values or modified *weighted* conformal p-values (Jin & Candès, 2023) for test data. Weighted p-values are particularly useful when the statistical assumption of exchangeability is weakened by covariate shift between calibration and test data. By providing these tools, `nonconform` enables researchers and practitioners to build anomaly detectors whose outputs are statistically controlled to cap the FDR at a desired nominal level:

The core assumption for the methods in `nonconform` is that the data is exchangeable, meaning the joint probability distribution is invariant to the order of observations. This makes the methods suitable for many cross-sectional data analysis tasks but not for time-series data where temporal ordering is informative.

Acknowledgements

This work was conducted in part within the research projects *Biflex Industrie* (grant number 01MV23020A) and *AutoDiagCM* (grant number 03EE2046B) funded by the *German Federal Ministry of Economic Affairs and Climate Action* (*BMWK*).

Bates, S., Candès, E., Lei, L., Romano, Y., & Sesia, M. (2023). Testing for outliers with conformal p-values. *The Annals of Statistics*, *51*(1). https://doi.org/10.1214/22-aos2244

Benjamini, Y., Heller, R., & Yekutieli, D. (2009). Selective inference in complex research. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1906), 4255–4271. https://doi.org/10.1098/rsta.2009.0127

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Chen, S., Qian, Z., Siu, W., Hu, X., Li, J., Li, S., Qin, Y., Yang, T., Xiao, Z., Ye, W., Zhang, Y., Dong, Y., & Zhao, Y. (2024). PyOD 2: A python library for outlier detection with LLM-powered model selection. *arXiv Preprint arXiv:2412.12154*.

Hennhofer, O., & Preisach, C. (2024).Leave-One-Out-, Bootstrap- and Cross-Conformal Anomaly Detectors . *2024 IEEE International Conference on Knowledge Graph (ICKG)*, 110–119. https://doi.org/10.1109/ICKG63256.2024.00022

Jin, Y., & Candès, E. J. (2023). *Model-free selective inference under covariate shift via weighted conformal p-values*. https://api.semanticscholar.org/CorpusID:259950903

Laxhammar, R., & Falkman, G. (2010). Conformal prediction for distribution-independent anomaly detection in streaming vessel data. *Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques*, 47–55. https://doi.org/10.1145/1833280.1833287

Lei, J., & Wasserman, L. (2013). Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *76*(1), 71–96. https://doi.org/10.1111/rssb.12021

Papadopoulos, H., Proedrou, K., Vovk, V., & Gammerman, A. (2002). Inductive confidence machines for regression. In *Machine learning: ECML 2002* (pp. 345–356). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-36755-1_29

Petsche, T., & Gluck, M. (1994). Workshop on novelty detection and adaptive system monitoring. *Advances in Neural Information Processing Systems (NIPS)*.

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world*. Springer-Verlag. ISBN: 0387001522

Zhao, Y., Nasrullah, Z., & Li, Z. (2019). PyOD: A python toolbox for scalable outlier detection. *Journal of Machine Learning Research*, *20*(96), 1–7. http://jmlr.org/papers/v20/19-011.html