The Journal of Open Source Software

nonconform: Conformal Anomaly Detection

Oliver Hennhöfer[1]

1 ISRG, Karlsruhe University of Applied Sciences

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

The requirement of uncertainty quantification for AI systems has become increasingly important. In the context of anomaly detection applications, this directly translates to controlling Type I (False Positive) error rates without compromising the statistical power of the applied detection procedure. Conformal Anomaly Detection (Laxhammar & Falkman, 2010) emerges as a promising approach for providing respective statistical guarantees by calibrating a given detector model. Instead of relying on anomaly scores and arbitrarily set thresholds, this approach converts the anomaly scores to statistically valid p-values that can then be adjusted by statistical methods that control the False Discovery Rate (FDR) (Benjamini & Hochberg, 1995) within a set of tested instances (Bates et al., 2023).

The Python library `nonconform` is an open-source software package that provides a range of tools to enable conformal inference (Lei & Wasserman, 2013; Papadopoulos et al., 2002; Vovk et al., 2005) for one-class classification (Petsche & Gluck, 1994). The library computes classical and weighted conformal p-values (Jin & Candès, 2023) using different conformalization strategies that make them suitable for application even in low-data regimes (Hennhofer & Preisach, 2024). The library integrates with the majority of `pyod` anomaly detection models (Chen et al., 2024; Zhao et al., 2019).

Statement of Need

The field of anomaly detection comprises methods for identifying observations that either deviate from the majority of observations or otherwise do not *conform* to an expected state of *normality*. The typical procedure leverages anomaly scores and thresholds to distinguish in-distribution data from out-of-distribution data. However, this approach does not provide statistical guarantees regarding its estimates. A major concern in anomaly detection is the rate of False Positives among proclaimed discoveries. Depending on the domain, False Positives can be expensive. Triggering *false alarms* too often results in *alert fatigue* and eventually renders the detection system ineffective and impractical.

In such contexts, it is necessary to control the proportion of False Positives relative to the entirety of proclaimed discoveries (the number of triggered alerts). In practice, this is measured by the FDR, which translates to:

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

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

Framing anomaly detection tasks as sets of statistical hypothesis tests, with H_0 claiming that the data is *normal* (no *discovery* to be made), enables controlling the FDR when statistically valid p-values (or test statistics) are available. When conducting multiple *simultaneous* hypothesis tests, it is furthermore necessary to *adjust* for multiple testing, as fixed *significance levels* (typically $\alpha \leq 0.05$) would lead to inflated overall error rates.

The `nonconform` (*non-conformity-based anomaly detection*) package provides the tools necessary for creating anomaly detectors whose outputs can be statistically controlled to cap the FDR at a nominal level among normal instances under exchangeability. It provides wrappers for a wide range of anomaly detectors (e.g., [Variational-]Autoencoder, IsolationForest, One-Class SVM) complemented by a rich range of conformalization strategies (mostly depending on the *data regime*) to compute classical conformal p-values or modified *weighted* conformal p-values. The need for *weighted* conformal p-values arises when the underlying statistical assumption of *exchangeability* is violated due to covariate shift between calibration and test data. Finally, `nonconform` offers built-in statistical adjustment measures like Benjamini-Hochberg (Benjamini & Hochberg, 1995) that correct obtained and statistically valid p-values for the multiple testing problem when testing a *batch* of observations simultaneously.

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