Research Mission - Lorenzo Perini, PhD

Research objective

Machine learning models excel at making predictions but are not always accurate. For instance, an assisted driving system failed to distinguish the white side of a trailer from the bright sky, causing a fatal accident[1]. Therefore, quantifying prediction uncertainty is increasingly important.

One such area is anomaly detection, where the goal is to identify examples that deviate from expected behavior [2], such as machine breakdowns [16]. Because these events are connected to high costs, trust in anomaly detectors is crucial. However, this is often complicated by the absence of labels, which makes anomaly detectors rely on strong data-driven intuitions about what constitutes an anomalous behavior to learn from data [3]. For example, anomalies in low-density regions can be detected by monitoring the negative log-likelihood of the data [1], while those with distinct feature values can be isolated by splitting the example space [6]. Unfortunately, anomalies can manifest in various ways, leading to high uncertainty near decision boundaries [9], which, in turn, reduces the user's trust in the system. That is, knowing upfront that the detector would make mispredictions for crucial events forces the user to double-check the data, which defeats the whole point of deploying an anomaly detector.

What if we could measure and exploit this uncertainty to develop a robust anomaly detector for real-world applications?

Some research directions

Uncertainty quantification. Uncertainty in anomaly detection includes aleatoric (data-related) and epistemic (model-related) uncertainties [5]. Quantifying these uncertainties is challenging due to the lack of labels, unrepresentative anomalies, and class imbalance. Key questions include: (a) *How can we quantify a detector's total uncertainty without labels?* (b) *Can we design specific estimators for each uncertainty type?*

Reducing uncertainty. Expert knowledge can improve predictions. For example, Active Learning [14], which targets examples with high model uncertainty, can help reduce epistemic uncertainty by labeling examples close to the model's decision boundary. However, reducing aleatoric and data-approximation uncertainties is less explored. For the former task, one can collect additional features or transform the existing ones. More importantly, for the latter one can generate examples by employing generative models, but the quality of generated examples must be measured to ensure they reduce uncertainty. Key questions are: (a) *How can we measure the quality of generated examples?* (b) *Can we quantify uncertainty reduction without domain knowledge?*

[1] https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF

Enhancing model reliability. Allowing models to abstain from predicting (i.e., saying "I don't know") when their uncertainty is too high can increase their reliability [4]. That is, whenever the model makes a prediction, the user knows that the uncertainty is low, and thus can rely on the output. This improves trust, as the user manually checks only the highly uncertain cases. Moreover, one can increase the quality of the model's output to an arbitrarily high value, by enlarging the set of examples where the model abstains. Key questions include: (a) *Can we guarantee high prediction accuracy with a given abstention rate?* (b) *What is the optimal abstention rate to balance prediction rate and accuracy?*

Prior Work

From anomaly scores to hard predictions. Anomaly detection involves assigning real-valued anomaly scores. Setting decision thresholds is challenging without labels. One option is to use the contamination factor (i.e., the proportion of anomalies), which is normally unknown. We pioneered estimating the contamination factor, addressing fully unsupervised settings [7], incorporating human-in-the-loop for targeted labeling [10], and transferring contamination factors across related assets [13].

Quantifying the uncertainty. Setting the decision threshold creates uncertainty: slightly perturbing the training data would change the decision threshold, which, in turn, could affect a test example's predicted class. Intuitively, this mainly affects the examples close to the decision boundary, and is connected to the concept of stability: the more stable a detector's output is for a test example, the less sensitive its predicted label is to changes in the training data. Our stability metric exactly captures this property of the detector's predictions [12].

Reducing uncertainty. Querying users for labels in practical scenarios is challenging. For example, in time series data, labels are often indicative of large time windows (e.g., a day), and learning from such labels is a hard task. We developed methods to learn from coarse-grained labels [11] and strategies to allocate labeling budgets across multiple assets [15].

Enhancing model reliability. Given a stability metric, one can decrease the uncertainty by introducing the human-in-the-loop. For example, by allowing the detector to abstain from unstable predictions, the user takes charge of labeling some test examples. We published the first survey on this field named Learning with Rejection [4] and pioneered the first approach that introduces a reject option into an unsupervised anomaly detector [8].

References

[1] Markus M Breunig et al. "LOF: identifying density-based local outliers". In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. 2000, pp. 93–104.

[2] Varun Chandola, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey". In: *ACM computing surveys (CSUR)* 41.3 (2009), p. 15.

[3] Songqiao Han et al. "Adbench: Anomaly detection benchmark". In: *Advances in Neural Information Processing Systems*. Vol. 35. 2022, pp. 32142–32159.

[4] Kilian Hendrickx et al. “Machine learning with a reject option: A survey”. In: *Machine Learning* (2024).

[5] Eyke Hüllermeier and Willem Waegeman. “Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods”. In: *Machine Learning* 110.3 (2021), pp. 457–506.

[6] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. “Isolation-based Anomaly Detection”. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012), pp. 1–39.

[7] Lorenzo Perini, Paul-Christian Bürkner, and Arto Klami. “Estimating the contamination factor’s distribution in unsupervised anomaly detection”. In: *International Conference on Machine Learning*. PMLR. 2023, pp. 27668–27679.

[8] Lorenzo Perini and Jesse Davis. “Unsupervised Anomaly Detection with Rejection”. In: *Advances in Neural Information Processing Systems*. 2023.

[9] Lorenzo Perini, Daniele Giannuzzi, and Jesse Davis. “How to Allocate your Label Budget? Choosing between Active Learning and Learning to Reject in Anomaly Detection”. In: *arXiv preprint arXiv:2301.02909* (2023).

[10] Lorenzo Perini, Vincent Vercruyssen, and Jesse Davis. “Class Prior Estimation in Active Positive and Unlabeled Learning.” In: *Proceedings of the 29th International Joint Conference on Artificial Intelligence and the 17th Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI)*. 2020.

[11] Lorenzo Perini, Vincent Vercruyssen, and Jesse Davis. “Learning from Positive and Unlabeled Multi-Instance Bags in Anomaly Detection”. In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023, pp. 1897–1906.

[12] Lorenzo Perini, Vincent Vercruyssen, and Jesse Davis. “Quantifying the confidence of anomaly detectors in their example-wise predictions”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 2020.

[13] Lorenzo Perini, Vincent Vercruyssen, and Jesse Davis. “Transferring the Contamination Factor between Anomaly Detection Domains by Shape Similarity”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 2022, pp. 4128–4136.

[14] Burr Settles. “Active learning”. In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6.1 (2012), pp. 1–114.

[15] Vincent Vercruyssen et al. “Multi-domain Active Learning for Semi-supervised Anomaly Detection”. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2022, pp. 485–501.

[16] Ling Xiang et al. “Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks”. In: *Applied Energy* 305 (2022), p. 117925.