**Abstract**

**Anomaly Transformer: Advancing Time-Series Anomaly Detection**

Time-series anomaly detection plays a critical role in ensuring the safety and efficiency of real-world systems such as industrial equipment, space exploration, and water treatment facilities. Identifying anomalies, which are rare and often concealed within vast amounts of normal data, is a complex challenge, particularly in unsupervised settings where labeled data is unavailable. Traditional approaches often fail to capture the intricate temporal dynamics and associations that are crucial for distinguishing anomalies from normal patterns. The **Anomaly Transformer** offers a breakthrough in addressing these limitations by introducing an innovative framework that leverages the power of Transformers to learn and exploit temporal associations.

The core contribution of the Anomaly Transformer is the **Anomaly-Attention mechanism**, a two-branch structure designed to model both **prior-associations** and **series-associations**. The prior-association focuses on adjacent temporal patterns, embodying the inductive bias that anomalies exhibit concentrated associations with nearby points due to continuity. The series-association, on the other hand, captures global temporal relationships across the entire time series. The divergence between these two associations, termed **Association Discrepancy**, serves as a robust anomaly criterion. Normal points tend to exhibit broader associations, while anomalies are characterized by concentrated, localized associations, making this measure inherently distinguishable.

A significant innovation within this framework is the use of a **minimax optimization strategy**. During the optimization process, the prior-association adapts to the temporal context by minimizing the association discrepancy, while the series-association maximizes the discrepancy to highlight abnormal points. This strategy enhances the model’s ability to separate normal and anomalous patterns, even in complex and noisy time-series data. Furthermore, the model integrates reconstruction loss to maintain the quality of temporal representations, ensuring a comprehensive anomaly detection criterion that combines reconstruction errors and association discrepancies.

The Anomaly Transformer has been extensively evaluated across six diverse benchmarks, encompassing applications in server monitoring, satellite telemetry, and critical infrastructure systems. It consistently achieves state-of-the-art performance, outperforming traditional methods such as density-estimation, clustering-based, and autoregression-based approaches. Notable datasets used in these evaluations include the Mars Science Laboratory (MSL) and Soil Moisture Active Passive (SMAP) from NASA, as well as the Secure Water Treatment (SWaT) dataset, all of which highlight the model’s robustness and versatility.

Key advantages of the Anomaly Transformer include its ability to:

1. Capture both local and global temporal dynamics through its dual-branch attention mechanism.
2. Leverage association discrepancies to define a novel, unsupervised anomaly criterion.
3. Enhance anomaly detection accuracy with a minimax strategy that amplifies the distinguishability of abnormal points.

This framework represents a significant step forward in time-series anomaly detection, addressing critical challenges in complex real-world systems. Its ability to generalize across applications underscores its potential for widespread adoption in domains where timely and accurate anomaly detection is paramount.