Roll: 1807065

Topic: Anomaly Detection

Data Anomaly Detection Based on Isolation Forest Algorithm: Liang Zhang & Lingyun Liu:

The initial step involves preparing the user access data stream by cleaning and organizing the data to ensure it is suitable for analysis. PCA is utilized to compress the high-dimensional data into a lower-dimensional space while preserving the most significant features. This step is crucial for enhancing the algorithm's efficiency by reducing the computational complexity. The core of the proposed method is the application of an Isolation Forest algorithm in a parallel processing framework. Isolation Forest works by isolating anomalies rather than profiling normal data points. It does so by randomly selecting features and then randomly selecting split values between the maximum and minimum values of these features. Anomalies are isolated closer to the root of the tree, thus having shorter path lengths. The optimized algorithm is rigorously tested using standard simulation datasets, such as Adult and NSL-KDD, to verify its effectiveness. The comparison with traditional Isolation Forest and LOF (Local Outlier Factor) algorithms demonstrates a significant improvement in both detection accuracy and processing speed.

An Abnormal Detection of Positive Active Total Power Based on Local Outlier Factor: Yan Liu, Ruiming Yuan & Sida Zheng:

The process begins with a preprocessing phase to clean the data, removing invalid values and employing Lagrangian interpolation to address missing data points. This mathematical technique reconstructs missing values based on the known points, ensuring the integrity and completeness of the dataset. Following this, the processed data is segmented according to the source equipment, recognizing that anomalies might differ across various types of equipment. The LOF algorithm is then applied to each segmented group. LOF works by calculating the local density deviation of a data point relative to its neighbors; a significantly lower density suggests an anomaly. This unsupervised method is particularly suited for this application due to its capacity to handle the inherent variability and complexity of PAP data.

Patch-based Sparse and Convolutional Autoencoders for Anomaly Detection in Hyperspectral Images: Amir Reza Rezvanian, Maryam Imani & Hassan Ghassemian:

Training on Background Data: The SAE is trained exclusively on samples considered to be "normal," which, in the case of HSIs, are the background pixels. This training process fine-tunes the network to capture the underlying structure and spectral signatures of the background.Sparsity-Induced Feature Selection: By enforcing sparsity, the SAE inherently performs a form of feature selection, prioritizing the most significant features in the background data. This is crucial in HSIs where the spectral dimensionality is high, and not all spectral bands contribute equally to distinguishing between normal and anomalous pixels. Anomaly Detection through Reconstruction Error: After training, the SAE is used to reconstruct both background and potential anomaly pixels. Anomalies are expected to result in a higher reconstruction error, as the SAE was not trained to capture their spectral signatures. Therefore, pixels with reconstruction errors exceeding a predefined threshold are flagged as anomalies.