MUltivariate Time Series ANomaly deTection

Framework (MUTANT)

Master Thesis

at Friedrich-Alexander-Universität Erlangen-Nürnberg
at the Masters in Data Science in the Data Science Department
Image Data Exploration and Analysis (IDEA) Lab

Principal Supervisor:	Prof. Dr. Bernhard Kainz
Associate Supervisor:	Mrs. Johanna Müller
Author:	MD FIROZUR RAHMAN Bahnstraße 91058 Erlangen +49 17658574913 firozur.rahman@fau.de 22975954

Submission:	22nd July 2024

The MUTANT framework, originally developed using Python, heavily relies on libraries such as NumPy, Torch, SciPy, Scikit-learn, and Pandas. However, the initial implementation had several issues:

(1) Redundant code and lack of descriptive comments to explain the purpose of various methods.

(2) Numerous library and syntax errors, particularly in the predict and reconstruct methods in Model.py and utils.py

(3) The model training and testing phases were not clearly separated.

(4) Slow data reading and feature generation processes without user notifications.

To address these issues, I made several improvements and successfully executed the code. The changes made are as follows:

(1) Fixed syntax errors, file size errors, label errors, library errors, and other types of errors.

(2) Provided comments throughout the code to explain the purpose and functionality of each method.

(3) Replaced certain loops and if-statements with default library functions to improve efficiency.

(4) Implemented a loading indicator to notify users when data is being read or features are being generated.

(5) Clearly divided the training and testing procedures post model generation.

(6) Developed a new method to return cluster labels.

(7) Created a mechanism to generate random group names and the number of data points in each group based on a self-supervised approach.

(8) Validation

(9) Reformatted the code for better readability and user experience

(10) Report and Output Customization for Clear User Understanding

Original repository: *https://github.com/Coac-syf/MUTANT*

Updated repository: *https://github.com/firozfau/anomaly-detection*

Short description of experiments and extensions:

To enhance the MUTANT framework, self-supervised learning is integrated to classify item types in addition to detecting anomalies. This is achieved by using clustering techniques like KMeans with PCA for dimensionality reduction. Future updates could include advanced clustering algorithms, additional self-supervised methods, and optimizations for handling larger datasets. .

0.1 Testing and training result.

```
(base) frzf7@frzf7 MUTANT % python main.py
Loading data of: SMAP
Data normalized
Train set shape: (135183, 25)
Test set shape: (427617, 25)
Test label shape: (427617,)
test_data: (277952, 25)
val_data: (149665, 25)

******************** Final result of Test data ******************************

Dataset-Name: SMAP
Threshold-value: 99.2500000000032
True Positive: 41805
False Positive: 1855
False Negative: 613
precision: 0.9575125971237946
recall: 0.9855485876313006
f_score: 0.9713233309601832
```

0.2 Testing result of Self-supervised learning

```
Performing brute-force search for the best threshold...
Performing clustering on test data...
Data reshaped for clustering: (5558640, 25)
Data shape: (5558640, 25)
Subsampling data to 10000 points.
Subsampled data shape: (10000, 25)
Applying PCA to reduce to 10 components.
Data shape after PCA: (10000, 10)
Number of clusters determined: 50
Starting KMeans fitting...
KMeans fitting completed.

Number of groups: 50
Group names and counts: {'Group0': 3971, ...... 'Group9': 625}

(base) frzf7@frzf7 MUTANT %
```