# Unsupervised Multivariate Time Series Clustering

**Md Monibor Rahman[1], L. Vidyaratne[2], A. Glandon[1], and K. Iftekharuddin[1]**

*[1]Vision Lab, Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA*

*[2] Jefferson Laboratory, Newport News, VA, USA*

## Summary

- **Shapelet learning is a process of discovering those Shapelets which contain the most informative features of the time series signal.**
- **This work proposes a generalized Shapelet learning method for unsupervised multivariate time series clustering.**
- **The proposed method is evaluated using an in-house multivariate time series dataset on detection of faults in the Jefferson Labs (JLab) Continuous Beam Accelerator Facility (CEBAF).**

## Background and Motivation

- **The CEBAF in the JLab has two linear accelerators that consist of 25 cryomodules each containing eight cavities.**
- **There is a very strong coupling between cavity to cavity. When one cavity trips off, the remaining seven cavities is likely to trip.**
- **The proposed generalized multivariate clustering approach is expected to develop a tool for automated fault identification and unsupervised data analytics for fault discovery.**

## Data set description

- **Each cryomodule in CEBAF has 8 cavities and every cavity generates 17 signals which are 8192 points long. Figure 1 presents the example of the input signals (cropped).**
- **The recorded data are arranged approximately 94% before the faults and 6% after the occurrence of the faults.**
- **We have used a total of 358 time series examples of three different types of fault to perform this analysis.**

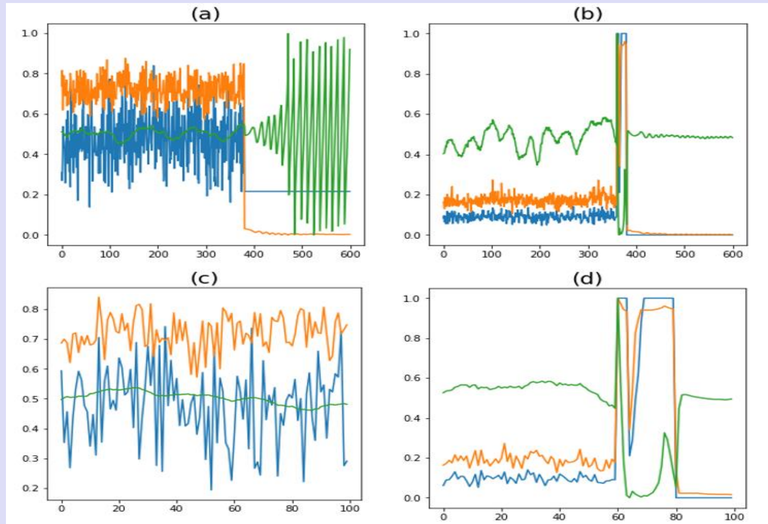


**Fig. 1: (a) and (b) represent two original multivariate time-series signals (cropped around faults) of length 600. (c) and (d) presents the multivariate Shapelet of length 100 which taken from the original time series fig. 1(a) and fig. 1(b) respectively.**

## Methodology

- **The distances of the multivariate shapelets and multivariate time series are calculated.**
- **The original multivariate time series data are mapped into the shapelet-based space.**
- **In shapelet-space, shapelet learning method learns the pseudo-class labels and a pseudo classifier using spectral analysis and regularized least-squares minimization [1].**
- **The multivariate shapelet is updated with the new learned pseudo-class labels and the pseudo classifier.**
- **The process repeats this until convergence.**

## Results

**The proposed method shows successful clustering performance of 3 faults (Quench 100ms, Equench and Microphonics) using 3 signals with average value of a precision of 0.732, recall of 0.717, F-score of 0.732, a rand index (RI) score of 0.812 and normalize mutual information (NMI) of 0.56 with overall less than 3% standard deviation in a five-fold cross validation evaluation. Figure 2 presents the 2D view of the clustering results. The method shown promising results for multiclass classification.**

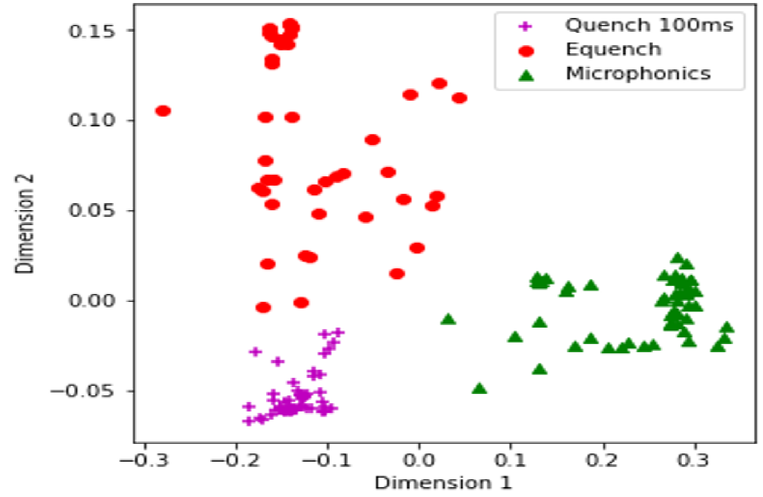


**Fig 2: Clustering results of the multivariate Shapelet learning method (2D view)**

**Reference**

**[1]** Qin Zhang et al "Salient Subsequence Learning for Time Series Clustering", IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 9, September 2019

**Contact:** mrahm006@odu.edu

**Acknowledgement**

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Contract No. DE-AC05-06OR23177. This work used the available resources in the high-performance computer (HPC) facility in the Wahab and Turing cluster of Old Dominion University, Norfolk, VA.