Project Title: Broad Advanced Intelligent Networked (BRAIN) System (SBIR/STTR: NASA)
Date: 09/25/2019 (Update 2)
Sub-contract: Dr. Pradeep Chowriappa, and Dr. Michael O'Neal in collaboration with AGNC.

Problem statement: Analyzing changes in evolving data from advanced habitation systems.

The goal of this task is to create a framework that monitors and provides for retrospective change detection in a multimodal system. **Our objective** is two-fold (a) To provide the mining of patterns from older data as changes in data could reflect long term/ previous trends. (b) Mining for patterns over near data – as recent changes in data could indicate the recurrences of the previously known pattern or an upcoming event.

Reported outcomes:

(a) An exhaustive survey of related datasets that capture multi-modal environments/systems.
(b) Implementation of a data streaming platform and using known techniques to capture concept drift (clustering) in evolving data.
(c) Implementation of change detection in time series datasets using python packages.
    a. Possible use of Spark SQL engine

Specific Aims worked on this week:

1. Searched for open source temporal datasets
    a. AMPds2 from Harvard University
2. Extract relevant data from these datasets
3. Using the statistical packages to analyze streams of data for the definition of micro and macro clusters over simulated streaming data.
    a. Simulated data streams with static structure, i.e. the statistical properties of the data stream do not change.
    b. Simulated data streams with concept drift, i.e. the statistical properties of the target variable, the model is trying to predict, changes with time in unforeseen ways.

Key Accomplishments:

1. Getting familiar with reading large files using R and Python panda and associated libraries
2. Trying to work with a dataset that seems to be similar to what was presented – by the TPOC
3. Investigated the known functions in R to build data stream framework:
    - Data Stream Data (DSD), simulates or connects to a data stream
    - Data Stream Task (DST), performs a stream mining task.

Future Work:

1. PCoE Datasets: (https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/)
    a. Work with the NASA bearing dataset (http://data-acoustics.com/measurements/bearing-faults/bearing-4/)
    b. Work with the IGBT accelerated aging dataset (https://c3.nasa.gov/dashlink/resources/134/)
2. To integrate the bearing dataset to the MOA-based techniques.
3. The implementation of uLSIF and its variant.

References:

[1] Makonin, Stephen. (2016). "AMPds2: The Almanac of Minutely Power dataset (Version 2)", https://doi.org/10.7910/DVN/FIE0S4, Harvard Dataverse, V2.

[2] Xiang Li, Wei Zhang, Qian Ding. (2018). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction.

[3] Hahsler, M., Bolanos, M., & Forrest, J. (2017). Introduction to stream: An Extensible Framework for Data Stream Clustering Research with R. *Journal of Statistical Software*, *76*(14), 1-50.

[4] Liu, Song & Yamada, Makoto & Collier, Nigel & Sugiyama, Masashi. (2012). Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation. Neural Networks. 10.1016/j.neunet.2013.01.012.

[5] Sugiyama, M. & Suzuki, T. & Kanamori, Takafumi. (2010). Density Ratio Estimation: A Comprehensive Review. RIMS Kokyuroku. 10-31.