Assignment 4 – Prognostics & Health Management

For an engineering system, Prognostics and Health Management (PHM) is a process aimed at f failure prevention and involving among other the estimation of Remaining Useful Life (RUL), so that the maintenance can be performed in a timely manner with a view to increasing the availability of the system. The signals acquired from an asset are continuously monitored, and corresponding data is processed in order to estimate the asset ‘s state of health.

To ensure optimal functioning of an asset, the PHM framework involves several steps: Anomaly Detection , Diagnosis, Prognosis, Health Management.

In this assignment, you will implement the three steps of PHM. In each step, the tasks necessary to achieve grade 3, grade 4 and grade 5 are described.

Part I . ANOMALY DETECTION i

You will be using a data-driven approach. You will use an open dataset (MNIST) available in PyTorch through torchvision. MNIST is an image dataset of 28x28 grayscale handwritten digits. The training set consists of 60K examples whereas test set is of 10K examples. You will be using 0-8 digit as normal data whereas digit 9 as an anomaly. After understanding the term anomaly detection, the basic implementation involves loading and pre-processing the data. Filtering the normal and anomalous data and later to visualize a sample from data loader. In an advanced task, you will train an autoencoder model and later evaluate the model on test image data.

The detailed break-up of the task list is given below.

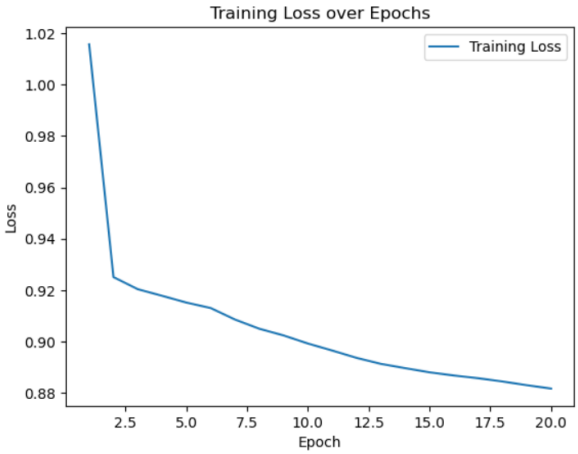
For implementation, you will use the Python programming language, pytorch framework, torchvision library for dataset. Along with this some other libraries from torch and basic libraries like numpy, matplotlib, etc.

Task to pass with grade 3

1. Understanding Anomaly Detection
    a. What is anomaly detection.
    b. Types of anomalies.
2. Load the dataset (open dataset MNIST using torchvision).
3. Transform the dataset.
    a. Normalize the data.
    b. Convert it to tensor.
4. Filter the normal data vs anomalous data.
    a. Use digit 0-8 as normal samples whereas digit 9 as anomaly.
5. Splitting the data in training and test sets.
    a. Train set will have digits 0-8 (normal data) whereas test set will have 0-9 (including anomalous data)
6. Check the dimension of train and test data loader.
    a. It should be batch size, channel, height, width.
7. Plotting one sample image from both train and test data loader.
8. Explain the steps to implement anomaly detection algorithm.

In a typical industry, where preventive maintenance is used, the monitored data is from a healthy asset. For anomaly detection when collected data is normal (non-anomalous), auto-encoder algorithm can be used to identify the images/data that deviate from the norm. Basically, an autoencoder is trained on normal images and will learn to reconstruct these images accurately. However, when an anomalous image is given as input, the reconstruction error will be higher. In this task you will train an autoencoder model for anomaly detection.

Sample training loss is shown in the figure below.



Task to pass with grade 4

1. Implement an Autoencoder model for anomaly detection.
    a. Use 3 linear layers with activation function for both encoder and decoder.
2. Choose appropriate hyper-parameters.
    a. Loss function.
    b. Optimizer.
    c. Learning rate.
    d. Number of epochs.
3. Plot the training loss.

After training the model, to evaluate you will use the test dataset. Compute an appropriate threshold value, for example, the mean + standard deviation of test losses. Use this threshold value to test on sample image from dataset to find whether a given image is an anomaly or not.

Task to pass with grade 5

1. Evaluate the model.
    a. Use test data to evaluate the model.
2. Calculate the mean and standard deviation of losses in test data.
3. Test the model with other sample images to find the given image as anomaly or not.
4. Build the confusion matrix.

Part II. DIAGNOSIS

Task to pass with grade 3

Try to discern different classes in the data set of Part I (the 8 different figures)

Task to pass with grade 4

Use the K-means clustering algorithm to sort out the data (you can use numpy and matplotlib)

Task to pass with grade 5

Evaluate the efficiency of the algorithm. Plot the loss function as a function of the number of clusters and determine the appropriate number of clusters.

Part III. PROGNOSIS

With the C-MAPPS data set of NASA (available on line) for turbofans, you will be using methods for RUL prediction.

Task to pass with grade 3

Download dataset and describe it

Task to pass with grade 4

With the FD01 data set of C-MAPPS, use the SVR (support vector regression) method to predict RUL.

Task to pass with grade 5

On the same dataset, use a DL (deep learning) method of your choice to perform the RUL prediction. Compare the result with that of the SVR method, by means of appropriate metrics.

REPORTING

You have to submit a 1-page report on this assignment. The report should contain:

1) Heading
2) Name of the participants
3) Problem description
4) Method of solution
    a. Information on the image dimension.
    b. Steps to implement anomaly detection.
5) Result figures
    a. Sample images.
    b. Loss plot.
    c. Sample test input and reconstructed image.
6) Your observations and reflections

The assignment will be considered complete, when the student has demonstrated the working code to the teacher, uploaded the code to GitHub repository, and submitted the report on canvas.