RLU estimation

Thai Nguyen

October 24, 2022

1 Why RUL estimation?

2 Mathematical formulation for RUL estimation problem

Suppose that we are provided with run-to-failure historical data of multiple machines of the same type. The historical data is supposed to cover a representative set of the considering machine type. The recored data of each machine ends when it reach a failure condition threshold. The objective is to learn a RUL model to predict the remaining lifetime of the considering machine type.

2.1 Training data

The training data is a set of the operation history of N machines of the same type denoted as $\mathcal{X} = \{X^n \mid n = 1, ..., N\}$. There are M measurements that quantify their health behavior during their operations (e.g. sensors that are installed on these machine to monitor their conditions). The data from the n^{th} machine throughout its useful lifetime produces a multivariate time series $X^n \in \mathbb{R}^{T^n \times M}$ in which T^n denotes the total number of time steps of machine n throughout is lifetime (in other words, T^n is the failure time of component n). We use the notation $X_t^n \in \mathbb{R}^M$ to denote the t^{th} timestamp of X^n where $t \in \{1, ..., T^n\}$. Indeed, X_t^n is a vector of M sensor values.

2.2 Testing data

The test set consists of historical data of K machines of the same type used in the training data denoted as $\mathcal{Z} = \left\{Z^k \mid k = 1, ..., K\right\}$ in which Z^k is a time series of k^{th} machine. The notation $Z_t^k \in \mathbb{R}^M$ is used to denote the k^{th} timestamp of Z^k where $t \in \left\{1, ..., L^k\right\}$ where L^k is the total number of time steps related to machine k in the test set. Obviously, the test set will not consist of all time steps up to the failure point, i.e.., L^k is generally smaller than the the failure time of component k denoted as $\bar{L}^k$.

We focus on estimating RUL of component k, $\bar{L}^k - L^k$, given the data from time step 1 to L^k. Note that $\bar{L}^k - L^k$ is also provided in the test set.

3 CMAPSS dataset

4 Data preparation

4.1 RUL target

We can generate the RUL for very time steps in a training trajectory X^n based on T^n. In the literature, there are two common models for generating RUL given the failure time, namely, linear and piece-wise linear model degradation model. These two models are mathematically presented in the following.

Linear model degradation model This kind of RUL model is very obvious considering the fact that we have the the failure point of each training trajectory (T^n). The RUL of machine n at time step t in the training set, R_t^n, is calculated as belows:

$$R_t^n = T^n - t \tag{1}$$

Piece-wise linear degradation model Since the degradation of a machine will generally not be noticeable unit it has been operating for some period of time. Therefore, it is probably reasonable to estimate RUL of a machine until it begins to degrade. For this reason, it seem to be ok to estimate the RUL when the machine is new as constant. As a result, the piece-wise linear degradation model is proposed to set an upper limit on the RUL target as belows:

$$R_t^n = \begin{cases} R_t^n = R^{max} & \text{if } t \leq T^{max} \\ R_t^n = T^n - t & \text{otherwise} \end{cases} \tag{2}$$

4.2 Data normalization