

Explanation of Machine learning application:

Accepts datasets of control inputs of process plant (multidimensional matrices per parameter e.g. observed parameter value and ideal parameter value (or range), location/target equipment/identifier and time in seconds) and generates a decision tree for process plant control outputs. Operators may use this suggestion for control output or change as per their understanding of process. The machine learning code uses this operator control values, and In loop/cycle again accepts next data set instance to process same, update decision tree, gain information. At some point during this training of the machine learning code, the code gains understanding of inter-dependencies among parameters and can model a mathematical functional dependency of a parameter on other set of parameters. A few parameters out of dataset (mainly their ideal parameter values), at each input dataset instance indicate the goal for process or the NORMAL scenario. The machine learning code keeps on processing the datasets at each time instance, and by predicting the inter dependence of plant input parameters on plant control output parameters, it can indicate if plant is heading towards its goal or away from it. The parameters which effectively control process are manipulated by expert operator to keep a few of observed parameters close to/within ideal values/range. Machine learning code predicts OFF-NORMAL conditions (only a few parameters have ideal set points, remaining need predictions for being OFF NORMAL). These predictions are qualified (YES/NO) by expert human operator during training. By this feedback and by the understanding of inter dependency among parameters, the machine learning code updates its decision tree continuously to correctly predict OFF-NORMAL conditions or desired values for plant control input parameters. Once this stage is achieved, the code can start to predict necessary actions (control output parameter values) in response to OFF-NORMAL /unknown process control input conditions excluding component failure. It predicts the operator action or sequence of it, so that a few observed parameters remain in range and process goal is maintained. A decision tree is statistical classification algorithm for dataset. If for example, dataset classification A results which is predicted to be OFF NORMAL, the decision tree has a set of values for control output or a sequence of gradual change in control output. The control output is predicted with goal to keep a few parameters at set-point and or return to NORMAL condition. If dataset classification is not A but B, then another set of output parameters get generated and so on. At each branch of decision tree, there is tag for being OFF-NORMAL/NORMAL. It generates rules for each control output parameter for each goal at different stages of process. For example a process can be broken into P stages. In P1, the goals are set for a subset of control input parameters. The control output is generated by machine learning code. Once the stage is successful, following operator's actions the process enters stage P2 and new decision tree is built for goals in stage P2 and so on.

Real life plant usage of such algorithm:

The laboratory plant setup operated by expert human to achieve set of goals is used as data generator for training data sets for machine learning code. The human operator's control actions following / anticipating unknown stimulation/random variations in environment in process state are part of the training data set. At each instance of processing training data set, the code learns about the interdependency among parameters. The code starts to learn from expert operator about his/her subconscious action-stimulus algorithm and generates a decision tree for each control output at process

stage for each process goal, a model in which control outputs should be generated so that a few plant set points are maintained in response to unknown process disturbances.

Below is explained why sample data is not required only in response to demand for sample data,

Generation of training input data sets and testing of machine learning code:

Process (control inputs) and expert human operator (control outputs) are required to be simulated for this contest. It is done by user defined multi variable equations with L.H.S equals to each control input parameter and R.H.S. to all other parameters and their assumed interdependencies. For example  $N_i = F_i\{M_m, N_n\}$ .  $i = 1,2,3\dots$ No of input parameters.  $m = 1,2,3\dots$ No of output parameters.  $n =$  all except that input parameter or less. So real example can be like this,  $N_1 = K_1.N_2 + K_2.d(N_3/N_2)/dT + \text{integral}\{K_3.N_1(T\text{-delay})\} + \dots + K_n.exp^{-N_n} + J_1.M_1 + \dots + J_m.M_m$ . Assuming that  $K$  and  $J$  are constants for each term and  $N, M$  are parameters. This way there will be Functions,  $F_i$  for each  $N$  (input) parameter depending on others. The equations can be generated using MS-Excel where each column will have a equation depending on other columns. These generated values need to be normalized in their respective equations by putting proper multipliers ( $10^{-n}$ ), so that the input parameters generated are within 0.0 and 1.0. There may be say 10000 entries generated for all parameters. At each entry there will be dataset inclusive of all  $N$  (input) and all  $M$  (output) parameters. The interdependency of parameters is not known to the machine learning code. It is known to the author of simulation which generated test dataset. Out of the 10000 entries, the machine learning code is trained with first say 8000 entries. We assume that the training is effective and machine learning code has understood the interdependency among parameters. The remaining 2000 entries excluding the values of control output ( $M$ ) are fed one by one. At each dataset entry to machine learning code, inclusive of only the control input ( $N$ ), the code may predict the values of control output ( $M$ ). In last a table is prepared for each output parameter ( $M$ ) comparing  $M$  values from simulation dataset and  $M$  values from machine learning code. The success of machine learning code contest is metric that most of the predictions are true.

Regards

Dhananjay G Parulekar,

TDD NRG BARC

02225591140(voice mail)