# Application of Neural Networks in oil refinery

## Building the Artificial Network

At the Tanura Oil Refinery the processes outputs are sent to a lab and the quality is measured. If the sample of the quality is not good enough the process needs to be "re-worked" resulting in costly expenses. The goal of applying a neural net to the process is to have the system learn the quality outcome based on the inputs. This will allow for the system to measure quality in a parallel way that can reduce the need for costly analyzers every time the outputs are sampled.

The use of soft computing methodologies; Neural Networks, Fuzzy Logic, Genetic Algorithms allow for more effective methods to help control highly nonlinear relationships. Neural Nets can be used for predictions by using mathematical coefficients that can be trained on historical data. The historical data is used to generalize the patterns that result in a desired outcome. The technique to train the neural net is called back-propagation. The data is split into training data for the model creation and test data to verify that the model is accurate enough to provide results given the known outcome. This helps to trust the model with unknown data in the real world.

Neural Nets require large data sets with a varied range to be able to accurately understand all the states it can come across during a real-world scenario. It is best that a many process variable as possible are fed into the dataset for the neural net to train on. The edge cases are best discovered by either running specific tests during plant operations to be able to capture the outcomes. The data would also need to be consistent. It is better for the data to be repeatable rather than accurate. An inaccurate sensor that measures a value consistently with an error is better than a sensor that varies its error in an unknown way. This type of noise is hard to train against since it cannot be abstracted from the true value. Systems like SCADA's or DCS's that capture sensor data and log it is more effective than handwritten logs for example. Sensors with flow or pressure can be averaged to remove noise.

All the input data should be collected in a reliable manner making sure that there is no lapse in the values and that they are all in sync so that the values are all captured at the same moment that the output is measured. A 3-month steady state data set was used in the experiment. The steady state of the data was found by making sure it was at least two time constant from the beginning of signal.

## Data Analysis

By coordinating a 'good' lab value result and the measured value results the ability for the model to accurately find what makes a given "lab" value becomes better. Eliminating bad values becomes a case of testing the data sets and discovering outliers. The method used in the paper was to utilize all 180 datasets except for 3 and rotate them through testing the three data sets and determining if it fell outside any repeatability tolerances found in the lab. The method identified the corrupt datasets.

Engineering knowledge is used to initially eliminate the process inputs but then the weights of the neural nets can be used to determine the what other process inputs can be removed. During the backpropagation training weights are associated to each of the inputs of the neural net. These weights can be used alongside engineering understanding so they can be prioritized. Albeit a lower weight does not necessarily mean the input is not needed.

The neural net can only be scored by how well it performs against a desired outcome. The outcomes that are desired in the project is the prediction of the Reid Vapor Pressure and the Naphtha 95% Cut-Point. These lab values are measured for 180 different input states. The input values are marked against the 180 samples taken. The idea would be to train a NN to provide the same lab measurement given the inputs. The results of the simulation show that there are many variables to consider when looking for its network performance.

The training for the Naphtha 95% cut-point consisted of 33 process variables with the sampled Naphtha 95% cut-point output. The dataset consisted of 70 for training and 15 for testing. A single hidden layer with 5 neurons was trained achieving an error of 0.01 in 10k iterations. To improve the model additional parameters where introduced and even an adaptive learning rate was added. These improvements resulting in a sum-squared error of 0.01 within only 3180 iterations. Attempts to further improve the outcome of the predictions resulting in adding hidden layers, and more neurons. Doing this resulting in memorization of the datasets and creating overfitting scenarios.

## Summary

**---**

## Introductions

- An increase in complexity of processes

- traditionally model based control systems are effective against local process changes in specifc range of operations

- Highly nonlinear relationship between inputs and outputs.

- Soft computing methodologies; Neural Networks, Fuzzy Logic, Genetic ALgorithms

- Samples of oil is taken twice a day and if it doesnt meet specification it is reprocessed resulting in cost and time.

- Online analyzers can be used to fine tune the process but a more cost effective ANN approach is proposed

- 3 Month data set to perform Inferential analysis of crude fractionation section of Ras Tanura Oil Rifinery at DHaraan

## Building the Artificial Network

- Neural Nets cna be used for predictions in processin plants

- Mathematical coefficeints for the ANN will be found using backpropogation training of the process inputs, until the "predicted" product quality and the "measured" product quality of processing output are equal.

- After training, data that was not used for training will be used for testing the models ability to predict the correct inputs to provide the desired output.

### Range of the Input Data

- training data should have the min.max ranges so that the neural net can learn all the edge cases

### Size of the Training data set

- A large data set is needed to average out inaccuracies int he dataset due to dynamic processes mistakes in the collection of the data.

### Aquiring the the training data set

- Least intrusive data collection method is during operation

- Training data needs to be varied to close gaps, captured duing forced plant tests.

-

### Validity of the Training Data Set

- industrial sensors; accuracy is not as important as repeatbaility

- it is best to take data from a SCADA or DCS system than oeprator logs

- process should be operating at steady state for at least two time constants

- flow and pressure should be averaged to remove noise

### Selecting Process ariables

- Process variables are inputs to the neural net

- It is best to choose as many input process variables as possible

- the neural ent will determine what is important during training

- get rid of least important inputs and add others if desired accuracy is not achievable after training.

## Data Analysis

- All input data mus tbe collected reliabliy

- Model parameters are built with a full data set

- A lab value that the output of the ann can be compared too must also be provided

- A complete dataset is all proces parameters and alab value

### Elimination of Bad Values

- Out of 180 datasets 3 datasets are set aside for testing.

- If the three tested data sets fall outside the repeatbility found in the lab, then its seen as deviant.

- this is done for all 180 datasets and all deviant datasets during testing are identified and ranked, removing the greatest outliers as corrupted datasets.

### Process Paramaters and thier efect on NN Prediction

- Use engineering knowledge of the process to elimnate unnecssary process inputs

- identify parameters based on the NN weights assigned , as long as its not a significant as an engineering parameter

- Adding parameters that enhance the engineering capability of the NN to understand effects better can improve the models performance

- It is important to align all the inputs so that they occur at the same time

### Implementation Procedures

1. Identify the application

2. Identify the Model Inputs

3. Define range of variables

4. Collect the plant data

5. Run Training program

6. model the results

- Is there enough variation in the data, get more plant tests to increase variation in the data.

- Is there not enough data? collect more plant data.

7. If model is ok then make it an Online application - yes send it to the online-dcs system if not then send it to the excel sheet.

- Identifying the application

The paper mentions that the application must be identified when creating a neural net. The neural nets constuction changes based on its purpose because it is trained for a specific outcome.

- Predictor Model tarining

The the training for the reid vapor pressure and for the Naphta 95% cutpoint requires a data set of all the process inputs and the desired outputs. The data is then scaled down to a workable range of 0-1 and the inputs are normalized before bieng applied to the NN. Matlab software is used to create and train a NN model. The meausrement used for scoring the model is the sum square error. The backpropogation training results in different outcomes based on the hyper parameters. Tuning it correctly allows the NN to converge to a minimum error.

Simulation Results and Discussions

The final goal is to remove the dependency of on on-line sample analyzers of the product quality

Naphta 95% cut Point

Conclusions