Machine Learning Based Modeling and Optimization of an Industrial Thermal Cracking Furnace

Melike Duvanoğlu,a,b,\* Gizem Kuşoğlu Kaya,a Onur Savran,a Erdal Aydın b,c

a Turkish Petroelum Refinery, Körfez, Kocaeli, 41790, Turkey

bDepartment of Chemical and Biological Engineering, Koç University, Rumelifeneri, İstanbul, 34450, Turkey

cKoç University, TUPRAS Energy Center (KUTEM), Rumelifeneri, İstanbul, 34450, Turkey

mduvanoglu22@ku.edu.tr, melike.duvanoglu@tupras.com.tr

Abstract

Machine learning methods can capture the distinctive characteristics of a system without any prior knowledge of the process given enough actual data. In addition, they are well suited to represent systems that are complicated for first principles modeling and have many unmeasured disturbances. Accordingly, data-based modeling for the thermal cracking furnace is a promising study using actual process data set and various machine learning methods. The study's focus is on the machine learning prediction of time-series Controlled Variables (CV), which is a prerequisite for using an Advanced Process Control (APC) system in a petrochemical plant. The most crucial component of an APC system is the prediction of the controlled variables and the adjustment of those anticipated values to bring them within the user's chosen range (Lee et al., 2023). Predicting the controlled variables is our main goal in this investigation. We specifically used a variety of machine learning approaches to forecast future controlled variables by utilizing historical controlled variables.

In this study, the cycle time of the furnace of a visbreaker unit and the temperature of the hottest zone of the furnace are modeled using different machine learning methods such as Support Vector Machines, Multiple Linear Regression, Decision Tree, Random Forest, and Artificial Neural Networks. Although the Random Forest model is good at predicting temperature and remaining day for the shut-down time, ANN model is used for process optimization purposes, by incorporating it in the fitness function of the genetic algorithm. When using a genetic algorithm (GA) to optimize a model for a specific task, the choice of the model as the fitness function is crucial. The fitness function evaluates how well a particular solution (set of model parameters or hyperparameters) performs the task at hand. The reason for using an Artificial Neural Network (ANN) as a fitness function in a genetic algorithm instead of a Random Forest (RF) is search space and differentiable nature of the ANN structure. Having estimated the cycle time by training the machine learning models, the inverse problem is attempted to solve such as calculating the optimal values of the features (controlled variables) for maximizing the operation time of the process within certain limits. This optimization problem is solved by sampling-based optimization methods formulating the trained machine learning models as fitness function. In this way, the necessary manipulated variables will be adjusted by the controller so that the unit can operate in the most efficient way.

**Keywords**: Visbreaker unit, thermal cracking furnace modelling, genetic algorithm, time-series prediction, controlled variables for advanced process control.

* 1. Introduction

Please use the SI set of units as much as possible. Wherever the application domain uses Thermal cracking process of heavy residue hydrocarbons occurs under severe thermal conditions. Visbreaking is a mild cracking in which mild heating is used to crack the residue, thereby reducing the viscosity while producing lighter products. While vacuum or atmospheric residue feedstock is heated and mildly cracked in the visbreaker furnace, soaker favors visbreaking between the furnace and the quenching step. Most of the limits faced by visbreaking units are caused by fouling. Thermal cracking of highly unstable and crucial asphaltenes causes fouling. When the temperature is reduced below thermal cracking by quenching after the furnace and soaker, this causes the production of coke particles at thermal cracking temperatures above 400–410°C (in the furnace) and subsequent precipitation. The coke is deposited on the walls of the furnace, which in turn leads to an increase in the tube wall temperatures and reduces the overall heat transfer coefficient. Determination of furnace fouling/coking tendency can help refiners boost visbreaker economics. (Speight, 1991).

A diagram of a chemical process

Description automatically generated

Figure 1. Simplified process flow diagram of Visbreaker process

The unit must be shut-down, and coke cleaning and maintenance must be carried out before the wall temperatures reach 600-650°C. Since many different types of crude oil-based residues are processed in the unit, the severity of the operation varies and this is the main reason for the coke formation process, which has already nonlinear nature. It is crucial to predict the degree of fouling or coking from the tube wall temperatures in different zones of the furnace and to optimize operation conditions while meeting product demand amount and desired quality.

In this study, various machine learning models are studied to predict the temperatures of the hottest part of the furnace and remaining day to shut down. ANN model is used as a fitness function, and the decision variables required to increase the time to shut down (extend the cycle time) and minimize the coil temperatures in the hottest region are determined by using genetic algorithm. Then these optimized controlled variables are set as a set point in the controller and the required action is taken in the controller by using manipulated variables.

* 1. Methodology
     1. Artificial Neural Network

An artificial neural network (ANN) is a mathematical model that attempts to simulate the structure and function of biological neural networks. The basic elements of all artificial neural networks are artificial neurons or simple mathematical models (functions). These patterns have three simple sets of rules: multiplication, addition, and activation. The inputs to the artificial neurons are weighted. That is, each input value is multiplied by its individual weight. At the heart of the artificial neuron is a summation function that adds all the weighted inputs and biases. At the output of an artificial neuron, the sum of the previously measured inputs and biases is passed through an activation function called a transfer function. (Van Laerhoven, 2000)

A feedforward ANN model is generally expressed as:

𝑦 = 𝑓1 (𝐴 ⋅ 𝑓2 (𝐵 ⋅ 𝑢 + 𝐶) + 𝐷) (1)

where *f1* and *f2* are output and hidden layer activation functions respectively. *A* and *B* are weight matrices; *C* and *D* are bias vectors; *u* is the input vector and *y* are the output vector. The dimensions of related ANN parameters are dependent on the number of inputs, outputs, and neurons, which are manually chosen prior to training.

* + 1. Genetic Algorithm

A genetic algorithm is defined as an optimization technique that is based on the principles of natural selection and is used to both constrained and unconstrained optimization situations. The idea of repeatedly changing the set of distinct answers forms the foundation of the algorithm. To create the next iteration and eventually arrive at an optimal solution, the algorithm randomly selects individuals at each stage. Genetic algorithms have been employed to forecast time series in the literature. One such example is the work of Khashei and Bijari, who demonstrated through examples how effective genetic algorithms can be for "time series modelling and forecasting. (Sohail, 2021)

*A diagram of a machine learning

Description automatically generated*

Figure 2. Schematic of the solution architecture

* 1. Results and Discussions

There are about 63 operational parameters of the visbreaker unit. Firstly, feature selection studies are carried out. Although the model is built using all features, modeling studies are developed by selecting only the most relevant features to reduce model complexity. From these data, 32 control parameters are selected that are related to the thermal cracking process specifications and provided information about the furnace operation and status. These inputs can be classified as heater outlet temperature, feed flow valve openings, flow temperature differences, coil flow differences, fuel gas pressure and temperature.

A graph showing a number of days

Description automatically generated

Figure 3. Target variable changes during the 10 years operation

These input and output variables are normalized between -1 and 1 values. Then, the ANN model is constructed by separating the 10-years of daily average data, that includes 2927 operation days and various cycles, 70% of training, 15% validation and 15% test data. The number of hidden layers is 16 and the *trainlm* function trains a feedforward neural network using the Levenberg-Marquardt optimization algorithm. The output of the ANN model is the coil 1 temperature of the hottest zone of the furnace.



Figure 4. Learning Curve for ANN

**Table 1.** ANN model results

|  |  |
| --- | --- |
| ***Performance*** | ***Mean Squared Error (MSE)*** |
| Best training | 3.43e-04 |
| Best validation | 8.68e-04 |
| Best test | 5.40e-04 |

Best validation performance observed in epoch number 14 as can be seen in Figure 4. It can be concluded that ANN model is successful for predicting the coil tube wall temperatures of the furnace from the performance metrics table. After the ANN model is built, this model is called by genetic algorithm as a fitness function that aims to minimize the temperature of the coil 1. The optimal decision variables that minimize the temperature of coil 1 is calculated. Although the 32 decision variable results exist, only 4 of them is shown in there.

A graph showing the number of days and days

Description automatically generated

Figure 5. Target variable and controlled variable changes 10 consecutive days.

Table 2.Some of the optimal solution results of ANN-GA hybrid model *(from 32 input variables only 4 are shown. Variables are normalized for proprietary reasons)*

|  |  |  |
| --- | --- | --- |
| ***Controlled Variable*** | ***Feature*** | ***Optimal Value*** |
| CV1 | Heater outlet temperature | -0.6066 |
| CV2 | Fuel gas pressure | -0.8045 |
| CV3 | Coil 1 valve opening | -0.8641 |
| CV4 | Coil 1 out temperature | -0.9146 |

Upon analyzing the optimum decision variables, the optimal decision is to reduce the 4 variables which are heater outlet temperature, fuel gas pressure, coil 1 valve opening, and coil 1 out temperature. From process perspective, it is logical to keep these variables to a minimum to reduce the coil 1tube wall temperature. To reduce the temperature in the furnace tube, it is required to reduce the heater outlet temperature, to reduce the pressure of the fuel gas, to send less feed to heater by reducing the valve opening of the flow of that tube, and to reduce the exit temperature of this coil at the end of the furnace.

For process, operation, and control engineers, changing these optimal control values with slower rates might be critical in terms of process safety and product specification and plant limits. Thus, it is planned to include rate change constraints into the formulation for hierarchical optimal operation and control. There are 32 control variables, and after they are sent as set points to the controller, the manipulated variables must be tracked by the DCS controllers. These control variables must be validated by working with safer and more accurate constraints so that the controller can take appropriate action.

* 1. Conclusions

In this study, we examined an industrial Visbreaker unit, which contains a thermal cracking furnace and has frequent shutdowns due to coking-related reasons. Since the coking and thermal cracking processes are quite complex, building physical models is challenging and time-consuming. Therefore, data-based modeling and optimization could be promising. It is concluded that the ANN model is successful in predicting the tube wall temperatures inside the furnace. Then, optimum decision variables, which will be sent to lower-level regulatory controllers, are computed to minimize the temperature. When the trends of these optimum (control) variables are examined, it is seen that they give reasonable and interpretable results. In the future stages, these control variables will be given as set points to the control system and the manipulated variable behavior will be examined and validated. By setting the control variables correctly and adjusting the appropriate manipulated variables of the controller, the cycle time of the furnace will be increased.

References

A. Sohail, 2021, Genetic Algorithms in the Fields of Artificial Intelligence and Data Sciences. Annals of Data Science, 10(4), 1007–1018.

J. Speight, 1991, The Chemistry and Technology of Petroleum, 2nd ed., Marcel Dekker, New York.

K. Van Laerhoven, 2000, Introduction to artificial neural networks. Proceedings Electronic Technology Directions.

M. Lee, Y. Yu, Y. Cheon, S. Baek, Y. Kim, K. Kim, H. Jung, D. Lim, H. Byun, C. Lee & J. Jeong, 2023, Machine Learning-Based Prediction of Controlled Variables of APC Systems Using Time-Series Data in the Petrochemical Industry. Processes, 11(7), 2091.