Published in final edited form as:

Ind Eng Chem Res. 2020 September 16; 59(37): 16357–16367. doi:10.1021/acs.iecr.0c02657.

Integrated Modeling of Transfer Learning and Intelligent Heuristic Optimization for Steam Cracking Process

Kexin Bi^{a,b}, Burcu Beykal^{c,d}, Styliani Avraamidou^d, Iosif Pappas^{c,d}, Efstratios N. Pistikopoulos^{c,d}, Tong Qiu^{a,b,*}

^aDepartment of Chemical Engineering, Tsinghua University, Beijing 100084, China

^bBeijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China

^cArtie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, 77840, United States

^dTexas A&M Energy Institute, Texas A&M University, College Station, TX, 77840, United States

Abstract

The construction and expansion of steam cracking plants and feedstock diversification have resulted in a significant demand for the numerical simulation and optimization of models to achieve molecular refining and intelligent manufacturing. However, the existing models cannot be widely applied in industrial practice because of the high computational expense, time-consumption, and data size requirements. In this paper, a high-performance optimization process, which integrates transfer learning and a heuristic algorithm, is proposed for the optimization of furnaces for various feedstocks. An effective transfer learning structure, based on motif feature of the reaction network, is designed and subsequent product distribution prediction program is compiled. Then a hybrid genetic algorithm and particle swarm optimization method is applied for the coil outlet temperature (COT) curve optimization using the derived prediction model, and the results are obtained for different pricing policies of products. The results are determined based on the weight coefficients of prices for different products, and could be further explained by the yield distribution pattern and reaction mechanism.

Keywords

steam cracking; transfer learning; coil outlet temperature curve; real-time optimization; reaction network

1. Introduction

Steam cracking of hydrocarbons is a main process for olefins production in the petrochemical industry. With the worldwide modernization and urbanization, the total demand for olefins is high and continuously increasing. As far as ethylene is concerned,

*Corresponding Author: qiuotong@tsinghua.edu.cn.

The authors declare no competing financial interest.

currently the demand is over 155 million tons per annum.¹ Driven by this increasing demand and an untapped opportunity for profitability, new ethylene plants are established and the existing ones are expanding their capacity.² Industry competition among these plants leads to the shortage and diversification of feedstock.³ Synchronously, enterprises are looking for more accessible lower priced feeds to face the challenge. The optimization process of the production in cracking furnace using these feeds is also essential for the most profitable and safest production process.⁴ With this goal in mind, there is a significant effort towards the process simulation and optimization in large numbers by petrochemical enterprises to achieve molecular refining and intelligent production.

Industrial software developers are helping these enterprises with the digital construction. In recent years, several state-of-the-art computational tools and software packages for the numerical simulation and optimization of steam cracking are developed including SPYRO,⁵ gPROMS,⁶ and CRACKER.⁷ However, these high-fidelity complex models are computationally expensive to be applied on various cracking combinations of feedstocks and furnaces from plants or to be embedded in the plant's entire real-time optimization (RTO) system. In such cases, representing a complex process with simplified data-driven mathematical models (i.e. surrogate models) has shown promise to achieve optimal or near-optimal solutions in a computationally efficient way.⁸ Specifically, considering the similarities between different feedstocks and reactors for steam cracking, machine learning methods have the most potential to be promoted in industrial practice.⁹

In this perspective, the previous study by Wang and Tang¹⁰ proposed a least-squares support vector machine (LSSVM) model to construct the mathematical relationship between operation parameters and yields of ethylene/propylene from steam cracking. Hough et al.¹¹ demonstrated that neural network and decision tree models were feasible approaches to represent high dimensional detailed kinetic models with high accuracy and reduced computational cost. Hua et al.¹² implemented feature mining based on motif detection from the reaction network, and optimized the convolutional neural network (CNN) structure to predict the yields of main products using the extracted features. The resulting model was shown to enhance the performance of numerical simulation while maintaining reaction features in the mechanism models. Nevertheless, there are conflicts between the high demands of data quantity (data-hungry) and low supplies from practical cases (data-lacking/data-poor) when applying these models. The data-driven models are data-hungry, and the plants are data-lacking where obtaining accurate and robust predictions for the steam cracking process is an important foundation to fulfill mandatory production targets and profit margins.

In addition to the lack of reliable real datasets for industrial applications of steam cracking reaction network modeling, a compatible optimization model is also required to achieve optimal operation of the production in furnaces. The main adjustable operation variables in steam cracking process are coil outlet temperature (COT), dilution ratio, and feed rate.¹³ Among these operation parameters, COT has a significant impact on product distribution. Trajectory of COT would not only influence the reaction selectivity, but it also would determine the coking process. The coke layer has a significant impact on the heat transfer of reactor, which is mainly decided by the coke thickness. Thus, deciding optimal trajectory of

COT over a given production length has attracted the interest of researchers. Keyvanloo et al.¹⁴ built a model of steam cracking based on the statistical design of experiments and obtained the optimization results from contour plots describing the response surface for main products. Geng et al.¹⁵ proposed an adaptive multi-objective particle swarm optimization (AMOPSO) algorithm based on the dynamic analytic hierarchy process (AHP) for the operation optimization of the ethylene cracking furnace. Zhang et al.¹⁶ implemented a parallel simulation on a multi-scale steam cracking model, which helped the plants increase the ethylene yield via multi-period operation optimization. However, the existing contributions in open literatures mainly focus on the optimization of the two main products, ethylene and propylene, and rarely describe a strategy for rapid model generation with new datasets (i.e., naphtha, topped oil, gas feedstock, and others) from plants.

To bridge the gap between data-hungry and data-poor entities while simultaneously identifying the optimal operating conditions for cracking furnaces with a range of product distributions, a two-phase integrated optimization framework is presented here, where: (1) Transfer learning techniques¹⁷ are introduced to numerically simulate the steam cracking process, and; (2) an intelligent heuristic optimization is proposed for the prediction and optimization of main products in the steam cracking process. The transfer learning process is an attractive methodology which requires a smaller dataset size than the usual machine learning model, once a basic model is established.¹⁸ For the steam cracking reaction model, as the source data from numerical simulation and the target data from plants are both labeled datasets, layer transfer technique¹⁹ is applied here, which belongs to parameter transfer technique²⁰. One of the CNN structures, mining features based on the motif proposed by Hua¹² is employed to explore the topology of reaction network and extract the relevant reaction kinetics for the steam cracking of naphtha. The trained, validated, and tested reaction network model then serves as the steam cracking problem simulator in the online optimization phase. In this phase, we focus on the COT curve optimization in each cycle of the furnaces using a hybrid genetic algorithm and particle swarm optimization (GA-PSO) framework,²¹ and quantify the effects of different sources of naphtha and various pricing policies on the product distribution using the reaction network analysis methods.

2. Methods

2.1 Integrated modeling and optimization of the steam cracking process

In a typical steam cracking process, a variety of hydrocarbons is usually used as feedstocks and first preheated in the convection section. The feedstocks are mixed with dilution steam and completely vaporized. Then, the mixture is sent to the radiation section for pyrolysis by the heat provided by fuel burners in furnace. After quenching, compression and separation, the products are collected for further production. Simultaneously, the coking process also occurs in the reactor, which makes the whole steam cracking process be carried out intermittently in cycles.

The model requirements of steam cracking plants are two-fold: (1) accurate prediction of the product distribution, and; (2) optimization of the production operation. To this end, an integrated framework (Figure 1) is proposed to return a packed model for each dataset provided by the plants. Our unified modeling and optimization framework is applicable to

varying types of feedstocks and furnaces due to the generalization of the transfer learning and intelligent heuristic optimization steps.

The framework starts with the offline modeling phase, where an industrial database and a simulation database are collected. These databases with feedstock properties, operation parameters, reactor conditions, and product yields are used to define the sub-models. These sub-models include the coke thickness model and the yield prediction model. Coefficients such as the weights of products and bound sets of decision variables,²² are also defined and solved during the offline modeling process. The sub-models and parameters are then implemented in the online optimization phase where GA-PSO algorithm is executed to obtain the optimal product distribution for a satisfactory operation. The details of each step of this framework are provided in the following sections.

2.2 COT curve design

In the cracking furnace, the heat flux of the tubular reactor could be manifested by COT. Thus, the COT curve has a decisive impact on the reaction and coking processes, which further determines the product distribution and energy consumption in the steam cracking production.²³ In industrial practice, the time-length between shut-downs of a furnace is set to be constant or within a small range of this constant to maintain the stability of cyclic production. Alternatively, many studies have shown that non-constant COT curves with varying shapes can yield higher profits.²⁴ However, obtaining highest profit usually needs extreme COT or coke thickness and frequent operation changes. The safety, stability, and other potential issues that are of extreme importance for operating industrial plants are neglected and should also be considered when deciding on the shape of COT curve.²⁵ In this work, a three-parameter COT curve is introduced for the simulation and optimization in a cycle of the steam cracking process, as shown in Figure 2.

For this application, the furnace running time-length is set as the abscissa scale of COT curve. After starting up, the furnace is first set under a starting COT (x_1) and then switched to an ending COT (x_2) on the operation switching date (x_3). The length of operation switching period (Δt) is determined by the difference of two COTs and the max increment of COT ($\Delta_m COT$):

$$\Delta t = \lceil |x_1 - x_2| / \Delta_m COT \rceil \quad (1)$$

The curve in switching period is generated by two values, including the difference of two COTs ($|x_1 - x_2|$) and the max increment of COT ($\Delta_m COT$). After the operation switching period (Δt) is calculated by Eq. (1), the increment absolute value of COT trajectories during switching period is generated as a vector $[\Delta_m COT, \Delta_m COT, \dots, |x_1 - x_2| - (\Delta t - 1) \times \Delta_m COT]$.

This proposed COT curve is a simplified approach inspired by previous research²⁶ and allows us to describe a key plant operation with only three parameters. Moreover, there are several advantages for this type of COT curve: (i) Enhances stability and prevents losses in

production by avoiding frequent operation switching;²⁷ (ii) Simplifies the subsequent optimization problem; and (iii) Enables the visualization and insight of reaction mechanism.

2.3 Offline modeling

2.3.1 Coking calculation—The reactor coking will also occur and influence the product distribution during steam cracking process.²⁸ The coke thickness is difficult to be characterized experimentally in industrial practice. However, Zhang¹⁶ proposed a surrogate coke thickness model based on the existing dataset, which can provide a fast and accurate prediction for parallel simulation, multi-period optimization, or machine learning. Thus, the regression formula, as shown in Equation (2), is applied here as a part of the neural network input to consider the coking effects.

$$\begin{aligned} \delta d_i &= d_{i+1} - d_i = a \times COT_i^2 + b \times i^2 + c \times COT_i \times i + d \times COT_i + e \times i + f \\ d_i &= \sum_{j=1}^i \delta d_j, i = 1, 2, \dots, N-1 \end{aligned} \quad (2)$$

In Equation (2), i is the period index; N the total amount of period in a cycle; d_i is the coke thickness in period i ; δd_i is the increment of coke thickness in period i ; a, b, c, d, e, f are the regression coefficients, which are re-calculated using the new datasets.

The growing coke inside the tube makes steam cracking a dynamic process. However, because of the slow coke formation rate, the entire cracking period could be divided into a series of quasi-stable state periods. Period is defined as the interval of adjacent operation marking dates, and in this work the marking dates are set as the $7n+3$ days ($n=0, 1, \dots, N-1$) of the entire cracking period. The coke thickness is updated between the adjacent period during simulation. Hence by adopting this technique we achieve the coke thickness model.

2.3.2 Transfer learning process for product distribution prediction—The previous steps of the framework provided necessary input data for numerical simulation of the steam cracking process, including feedstock properties, furnace geometry, operation condition, and coke thickness. As the goal is to use this data in the transfer learning model based on reaction mechanism, a hybrid mechanism feature and data-driven neural network structure of steam cracking prediction is designed and shown in Figure 3. The reaction network part is implemented first for feature extraction from reaction network, and the features are further analyzed and regressed in the following product prediction part together with the operation conditions and coke thickness curve.

In the reaction network part, the input is a motif-feature matrix, which represents the feedstock composition and reaction network topology, and the output is extracted reaction features of specific combination of the feedstock and reactor. The motif-feature matrix is generated using the method proposed by Hua by the following steps. Firstly, the feedstock properties are converted into molecular compositions using the molecular reconstruction model in Ethylene cracker Simulation and Optimization System (EcSOS).²⁹ Then, motif detection techniques are applied here for topology pattern extraction from the reaction network,^{30–31} and the extracted motifs are vectorized and stacked as a matrix. Finally, the

molecular composition values are filled into corresponding positions of the matrix, so that the motif-feature matrix combines information of the feedstock compositions and the reaction network topology. Taking naphtha as an example, a motif-featured matrix with a size of 34942×3 is collected as the input of reaction network part of the neural network structure, and the GoogLeNet structure³² is applied here for dimension reduction and further feature extraction from the matrix to enhance the generalization ability.

The final reaction features, integrated with operating conditions and coke thickness curve as additional inputs, are entered into product prediction part of the neural network structure. The operating conditions include the COT shape variables (x_1 , x_2 , x_3), dilution ratio, feed rate, coil inlet temperature, coil inlet pressure, coil outlet pressure, and calculated coke thickness curve includes the thickness on $7n+3$ days ($n=0, 1, \dots, N-1$). Here, fully connected layers are added to enhance the learning ability for a specific type of furnace. The model outputs include the product distribution and restriction variable of the production to ensure process safety. For example, the tube maximum temperature (TMT) could be a restriction to prevent the tube reactor from industrial faults.

For the cracking process using similar types of furnace and different types of feedstock, there are three reasons why the transfer learning method can be applied: (i) high resemblance of reaction networks; (ii) similar reactor models; (iii) similar main product evolving trends. Hence, with enough simulation data and several industrial data for a certain combination of feedstock and reactor, the transfer learning process could be implemented using a small quantity of industrial data of another combination. As the datasets used during learning process are all labeled, layer transfer technique¹⁹ is employed here. A schematic diagram of layer transfer using in the proposed steam cracking neural network is shown in Figure 4.

As is shown in Figure 4, the transfer learning process was implemented on two datasets. In order to obtain reaction network features in the mechanism model and transfer them to a new model, the simulation results in EcSOS software were collected as the source data and the industrial calibration results in real plants were collected as the target data. After the training of original model shown in Figure 3 using the source data, the layer structure was directly transplanted to the new model. The parameters in reaction network part of the neural network structure were directly copied in the new model and fixed when re-training. While in the production prediction part, the parameters were set as initial values in the corresponding part of new model and re-trained using the target data. The features of reaction network were kept intact because the processing procedures from the feedstock composition to the reaction features were exactly the same for the original model and the new model. The diversity of furnaces and operational parameters was embodied by the retraining of production prediction part.

2.4 Online optimization

After the data collection and transfer learning model training steps, a reaction mechanism-based product distribution prediction results could be obtained promptly. The intelligent and high-performance prediction laid a solid foundation for the operation optimization of steam cracking process.³³ Multiple types of products are generated during the production of

furnace, and a new optimization model is built to consider these products comprehensively. In this model, a set of weights related to the price and production requirement is assigned to the main products for operation optimization. There are also bounds for the decision variables which are also the operation parameters of COT curve. Safety-related variables, TMT in multi-period of the production,³⁴ are also considered here and set as an output of the transfer learning model. The mathematical formulation of the corresponding optimization problem is given as:

$$\begin{aligned}
 & \max_{x_1, x_2, x_3} \quad w_1 \cdot y_1 + w_2 \cdot y_2 + \dots + w_n \cdot y_n \\
 & s.t. \\
 & y_1 = f_1(x_1, x_2, x_3, motif\ matrix, coke\ thickness, operation\ conditions) \\
 & y_2 = f_2(x_1, x_2, x_3, motif\ matrix, coke\ thickness, operation\ conditions) \\
 & \vdots \\
 & y_n = f_n(x_1, x_2, x_3, motif\ matrix, coke\ thickness, operation\ conditions) \\
 & TMT = f_{n+1}(x_1, x_2, x_3, motif\ matrix, coke\ thickness, operation\ conditions) \\
 & lb_1 \leq x_1 \leq ub_1 \\
 & lb_2 \leq x_2 \leq ub_2 \\
 & lb_3 \leq x_3 \leq ub_3 \\
 & TMT \leq ub_4
 \end{aligned} \tag{3}$$

where n is the amount of all types of products; $w_1, w_2, \dots, w_n$ are the set of weights; $y_1, y_2, \dots, y_n$ are the yields of products; $f_1, f_2, \dots, f_{n+1}$ are the generated prediction models for product yields and TMT using the transfer learning techniques mentioned above; x_1, x_2, x_3 are the parameters in the constructed COT curve to be optimized; $lb_1, lb_2, lb_3, ub_1, ub_2, ub_3, ub_4$ are the lower and upper bound for the decision variables.

After compiling the transfer learning models for product distribution prediction, the optimization phase calls these functions during the execution of the optimization. Heuristic optimization methods are widely used in the optimal process design and operation problems in steam cracking process, especially the approaches based on intelligent population.³⁵ In previous studies, GA, PSO and their variants are adopted in most of the optimization problems for cracking furnaces.^{36–38} From the perspective of industrial application, hybrid GA-PSO is mature and well-proven by our previous researches. It is also compiled and applied as a plug-in for industrial internet for petrochemical enterprises. Therefore, in this work, a hybrid GA-PSO algorithm²¹ is employed to solve this nonlinear optimization problem. The detailed setting of the GA-PSO algorithm is shown in Table 1.

To solve this problem, we set the termination criteria as the fitness value of best individual(s) obtained 10 times changed within a small range ($\epsilon < 0.0001$). It is a common termination criterion used in GA, PSO and their variant to ensure the global optimal solution.

2.5 Industrial cases and datasets

An olefin plant in the northwestern part of China is selected as an industrial case study and the schematic diagram of millisecond furnace operating in the plant is shown in Figure 5. The feedstock properties, operation parameters, and the product distribution data are collected from real process. To further expand the dataset size for the proposed machine learning analysis, real industrial data are supplemented with simulated data. The simulation

data were collected from EcSOS software, which is jointly developed by PetroChina and the Department of Chemical Engineering at Tsinghua University.

The dimensionality of process design space is related to the operational inputs of production prediction part, which include the COT shape variables (x_1 , x_2 , x_3), dilution ratio, feed rate, coil inlet temperature, coil inlet pressure, coil outlet pressure. When spanning this space, orthogonal experimental design was applied and the variables were divided into two groups. The first group was composed of COT shape variables (x_1 , x_2 , x_3), dilution ratio, which are often used for product distribution adjustment. The second group was composed of feed rate, coil inlet temperature, coil inlet pressure, coil outlet pressure. The yields of ethylene, propylene and butadiene were the most concerned in all products, therefore the L9 orthogonal array was adopted as the basic experimental design unit. For each variable in the first group, 9 values at equal interval were selected for experiment and adjacent 3 values were divided into a subspace. For each variable in the second group, 3 values at equal interval were selected for experiment. The entire 8-dimensional design space was first divided into $3^4=81$ subspaces using the 4 variables in the first group. Then for each subspace, two L9 orthogonal arrays were designed for two variable groups, and the quantity of total circumstances were $9 \times 9=81$. Then a total of 6561 experimental locations of samples were selected for simulation. After the sample location generation, coupled simulations of recirculation zonal firebox model and detailed kinetic reactor model using EcSOS software packages were implemented using the local naphtha and millisecond furnace.³⁹ 4593 samples (70%) were divided into training set and the others were divided into testing set randomly using the 'train_test_split' algorithm from sklearn function library. When training the basic model, 10-fold cross validation was applied in order to make a full use of the dataset.

For transfer learning process, the simulation results were then collected and packed as source data. The target data were composed of 21 sets of real industrial data with another type of naphtha and reactor. Parallel samples and experimental calibration strategies were adopted to guarantee that the quality of industrial data is good enough for modeling. According to the design data and production experience of the reactor, the total production length is set as 45 days and the max increment of COT (ΔCOT) is set as 2 °C.

3. Results and discussion

3.1 Optimization results under different pricing policies

The prices of products are continuing to vary, and the pricing policy will have a great impact on the production strategy.⁴⁰ The investigated plants use the directly proportional function as the conversion between prices and weights, and set the weight of ethylene as 1.00. Here we collected three sets of historical weights for products, which are listed in Table 2.

After entering the bounds and transfer learning functions, the GA-PSO optimization was implemented with different sets of weights. The COT curve results using different pricing policies are shown in Figure 6, and the product distribution results are plotted in Figure 7. A base case using constant COT is used for result comparison.

For Set 1, the weight of ethylene was relatively high among all products, thus ethylene could be regarded as the main product. Figure 6 and Figure 7 showed that the production process has a higher temperature preference to produce more ethylene at this time, and the ending COT (x_2) reached the industrial limit of the millisecond furnace. For Set 2, propylene and butadiene were the main products, and a lower COT curve was obtained to adjust the product distribution for higher propylene and butadiene yields. For Set 3, the homologs of benzene were considered more important, and a lower starting COT (x_1) was required for a reasonable product distribution with higher benzene yields. The objective function values of the optimized product distribution versus the base case is listed in Table 3 to show the improvement of the profit margin after optimization.

3.2 Yield pattern analysis using a reaction network characterization tool

To better display the changing pattern of the objective function (OBJ) with the temperature variable, we set the operation switching date (x_3) equaling to the 23rd day of a cycle, and plotted the $OBJ - x_1 - x_2$ curve in Figure 8.

From the manifold in Figure 8 (a), we could also conclude that the objective function value would increase with increasing temperature when ethylene was the main product. However, the manifolds in Figure 8 (b) and Figure 8 (c) were not monotonic, and the optimal points of the COTs were lower for a profitable product distribution compared to Figure 8 (a). The optimization results in Figure 6 and changing pattern in Figure 8 could be mutually confirmed.

A reaction network characterization tool using the sub-network reconstruction method proposed in our previous work could give a more comprehensive explanation of the optimization results,⁴¹ and show how the COT curve influences the results from the molecular/free radical perspective. Inspired by PageRank algorithm which could be applied for page level sorting or social network analysis, the reaction network could be abstracted into a 'social network' that the nodes are connected using mass flow. To analyze the importance of the substances and search for the key interaction evolving trends in the entire reaction network, a sub-network using substances as nodes and reaction rates as connection weights is reconstructed and re-iterated subsequently using PageRank algorithm. The re-iterated PageRank value (RPRV) could be considered as the substance node importance in accordance with the significance of PageRank algorithm, and the key interaction evolving trends of a specific node could be backtracked from the reaction network.

Here, we took the results of Set 2 as an example to explain the curve shape in Figure 6 and Figure 8. First, the cracking furnace model was simulated using mechanism model of EcSOS under different COTs. After extracting the connections and rate information from the reaction network, we could obtain the relationship of specific substance node importance, RPRV and the production yield (PY) using sub-network redistribution, and the node information of propylene and butadiene was shown in Figure 9.

As shown in Figure 9, the node importance index (RPRV) that is calculated by sub-network reconstruction characterization is continuously increasing with the increasing COT. However, there is an inflection point in the production yield (PY) curve. An inference could

be drawn from the significance of RPRV, that is, the RPRV can be considered as mass quantity accumulated in a specific substance node, thus the generation of propylene and butadiene increased. However, if the COT reached the level of inflection point in Figure 9, the tendency of propylene and butadiene participating in other reactions would be extended. This will change the role of propylene and butadiene from resultant to mid-product in the entire reaction network with the gradual increase of COT. The inference is also validated by backtracking the related reactions in the entire network. Here we selected several example reactions related to the generation and consumption of propylene and butadiene for the mechanism schematic. The reaction rate fluctuation trends with the increasing COT are listed in Table 4, and related reaction routes are plotted in Figure 10.

As shown in Table 4 and Figure 10, the decomposition rate of hydrocarbon had a positive correlation with the COT, and the generation rate of propylene and butadiene from this type of reaction would also slightly increase. However, when the COT reached the level of yield reflection point, the reactions of propylene and butadiene converting into ethylene and other lighter products (i.e., methane) are strongly accelerated, which hinders the mass accumulation of propylene and butadiene in the whole reaction network and causes a decrease in the product yield. This explains why the yield of ethylene is continuously increasing while the yields of propylene and butadiene have an inflection point with respect to the COT. The COT range of optimal operation curve for Set 2 in Figure 6 was from 863.0 °C to 865.9 °C, which also coincided with the yield inflection COT range of propylene and butadiene yield. Hence, the optimization results could be explained by yield distribution pattern and reaction mechanisms, which provided mutual validation among these methods.

3.3 Computational Performance

For the numerical simulation and process optimization, the performance of model is an essential criterion for its potential application in industrial practice. The mechanism model for product prediction of steam cracking will approximately take from 30 seconds to 20 minutes using existing software packages. In addition, a new model needs to be readjusted when using a new type of feedstock or reactor. For the transfer learning method, the re-training of datasets for new models would be much faster, and the detailed performance was listed in Table 5.

In the training of transfer learning model, Percentage Mean Absolute Deviation function is employed as a loss function (L), which could be expressed as:

$$L = \frac{100}{N} \sum_{i=1}^N \frac{|y^{(n)} - o^{(n)}|}{y^{(n)}} \quad (4)$$

A prediction comparison of the transfer learning model and the EcSOS was also made to display the accuracy explicitly. The yields of 9 main products during the production were employed as indexes, which is shown in Table 6.

For various COT curves obtained from different weight sets, the relative predicted errors of transfer learning model are all within in the range of $\pm 5\%$, which is acceptable for industrial use.

The whole machine learning process for product distribution prediction of the steam cracking production was programmed on Tensorflow backend⁴² by Python 3.6. Stochastic gradient descent algorithm (SGD)⁴³ was selected as the optimization method for parameter training. After statistics, there were 0.35 million parameters in the whole network, which needed to be first trained in the basic model. However, in training tasks of transfer learning, only 0.14 million parameters were re-trained in fully connected operations, and the convolution operations in the reaction network part were completely stored. Thus, the transfer tasks were much less time-consuming.

Furthermore, a performance comparison of a PC with a CPU of i7-4770 @3.40GHz, a work station with a CPU of i7-9750 @4.50GHz and a GPU of Nvidia RTX 2070mq (8GB), a server with a CPU of Xeon Phi 5110P @2.50GHz and a GPU of Tesla V100 (32GB) on High-Performance Research Computing (HPRC) of Texas A&M was made to make a recommendation for an enterprise to select the computing facility. Here, we can see that the accuracy of machine learning process, which was quantified by a loss function, is similar among different platforms. The loss validation on testing set also prove the extendibility of the proposed model on these platforms. However, the time-consumption is quite different, especially between the training on CPU and GPU. For a newly built plant with great modeling needs, a high-performance computing cluster is recommended for leasing to complete the training of various prediction models. For a plant with needs for model updating, a workstation with a decent GPU is recommended to build for saving costs.

After training of machine learning model, the prediction program was compiled and ran in the optimization model. The prediction process could be completed within ~ 0.1 seconds per time, which met the demand of multiple calls during the GA-PSO optimization process. After testing, the GA-PSO algorithm could usually complete the optimization within 100 seconds and find the global solution, which was acceptable for industrial RTO. Detailed performance of the GA-PSO optimization process is listed in Table 7.

4. Conclusions

In this work, a high-performance optimization formulation for steam cracking production is proposed and evaluated. The transfer learning method, which was applied as the product distribution prediction model, not only provided an approach for quick modeling and prediction, but also incorporated the reaction mechanisms. Further explanations using the reaction network characterization tool also demonstrated the accuracy of optimization results. The compiled programs have the potential to be part of a new software package for steam cracking plants.

In addition, the whole process can also be applied to more types of RTO tasks. For model-hungry plants without molecular modeling experience, an integrated model with a transfer

learning process, based on the existing mechanism method and an intelligent optimization algorithm, could make the decision making more beneficial and profitable.

Acknowledgements

The authors gratefully acknowledge the financial support received from the National Natural Science Foundation of China (Grant No. U1462206, 21991100, 21991104), the U.S. National Institutes of Health Superfund Research Program (NIH P42-ES027704), the National Science Foundation (Grant No. 1705423), and the Texas A&M Energy Institute. Portions of this research were conducted with the advanced computing resources provided by Texas A&M High-Performance Research Computing. The manuscript contents are solely the responsibility of the grantee and do not necessarily represent the official views of the NIH. Further, NIH does not endorse the purchase of any commercial products or services mentioned in the publication.

References

- (1). Sadrameli SM Thermal / Catalytic Cracking of Hydrocarbons for the Production of Olefins : A State-of-the-Art Review I : Thermal Cracking Review. *Fuel* 2015, 140, 102–115.
- (2). EIA. U.S. Ethane Consumption, Exports to Increase as New Petrochemical Plants Come Online. 2018, <https://www.eia.gov/todayinenergy/detail.php?id=35012>.
- (3). Chen S Analysis of Current Status of Steam Cracking Feed Production and Measures for Maximization of Steam Cracking Feed. *China Pet. Process Pe.* 2005, 2, 41–46.
- (4). Jin Y; Li J; Du W; Qian F Integrated Operation and Cyclic Scheduling Optimization for an Ethylene Cracking Furnaces System. *Ind. Eng. Chem. Res.* 2015, 54 (15), 3844–3854.
- (5). Goethem M. W. M. Van; Kleinendorst FI; Leeuwen C. Van; Velzen N. Van. Equation-Based SPYRO ® Model and Solver for the Simulation of the Steam Cracking Process. *Chem. Eng. Res. Des.* 2001, 25, 905–911.
- (6). Haghighi SS; Rahimpour MR; Raeissi S; Dehghani O Investigation of ethylene production in naphtha thermal cracking plant in presence of steam and carbon dioxide. *Chem. Eng. J.* 2013, 228, 1158–1167.
- (7). Joo E; Lee K; Lee M; Park S CRACKER — a PC Based Simulator for Industrial Cracking Furnaces. *Comput. Chem. Eng.* 2000, 24 (2), 1523–1528.
- (8). Beykal B; Boukouvala F; Floudas CA; Sorek N; Zalavadia H; Gildin E Global Optimization of Grey-Box Computational Systems Using Surrogate Functions and Application to Highly Constrained Oil-Field Operations. *Comput. Chem. Eng.* 2018, 114, 99–110.
- (9). Fakhroleslam M; Sadrameli SM Thermal / Catalytic Cracking of Hydrocarbons for the Production of Olefins; a State-of-the-Art Review III : Process Modeling and Simulation. *Fuel* 2019, 252 (April), 553–566.
- (10). Wang X; Tang L Multiobjective Operation Optimization of Naphtha Pyrolysis Process Using Parallel Differential Evolution. *Ind. Eng. Chem. Res.* 2013, 52 (40), 14415–14428.
- (11). Hough BR; Beck DAC; Schwartz DT; Pfaendner J Application of Machine Learning to Pyrolysis Reaction Networks : Reducing Model Solution Time to Enable Process Optimization. *Comput. Chem. Eng.* 2017, 104, 56–63.
- (12). Hua F; Fang Z; Qiu T Application of Convolutional Neural Networks to Large-Scale Naphtha Pyrolysis Kinetic Modeling. *Chinese J. Chem. Eng.* 2018, 26 (12), 2562–2572.
- (13). Zhang Y; Hu G; Du W; Qian F Chapter Three - Steam Cracking and EDC Furnace Simulation In *Advances in Chemical Engineering: Thermochemical Process Engineering*; Van Geem K Eds.; Academic Press, 2016; Vol. 49, pp 199–272.
- (14). Keyvanloo K; Towfighi J; Sadrameli SM; Mohamadalizadeh A Investigating the Effect of Key Factors, Their Interactions and Optimization of Naphtha Steam Cracking by Statistical Design of Experiments. *J. Anal. Appl. Pyrolysis* 2010, 87 (2), 224–230.
- (15). Geng Z; Wang Z; Zhu Q; Han Y Multi-Objective Operation Optimization of Ethylene Cracking Furnace Based on AMOPSO Algorithm. *Chem. Eng. Sci.* 2016, 153, 21–33.
- (16). Zhang L; Wang H; Qiu T; Chen B Simulation and Optimization of Multi-Period Steam Cracking Process; Patel V Ed.; IntechOpen: Rijeka, 2015; pp 117–138.

- (17). Pan SJ; Yang Q A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* 2010, 22 (10), 1345–1359.
- (18). Shin H; Roth HR; Gao M; Lu L; Xu Z; Nogues I; Yao J; Mollura D; Summers RM Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* 2016, 35 (5), 1285–1298. [PubMed: 26886976]
- (19). Yosinski J; Clune J; Bengio Y; Lipson H How Transferable Are Features in Deep Neural Networks? In *Advances in Neural Information Processing Systems 27*; Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, Eds.; Curran Associates, Inc., 2014; pp 3320–3328.
- (20). Zhou S; Du W Modeling of ethylene cracking furnace yields based on transfer learning. *CIESC J.* 2016, 65(12), 492–498.
- (21). Bi K; Qiu T An Intelligent SVM Modeling Process for Crude Oil Properties Prediction Based on a Hybrid GA-PSO Method. *Chinese J. Chem. Eng.* 2019, 27 (8), 1888–1894.
- (22). Karimzadeh R; Godini HR; Ghashghaee M Flowsheeting of Steam Cracking Furnaces. *Chem. Eng. Res. Des.* 2009, 87 (1), 36–46.
- (23). Yu K; While L; Reynolds M; Wang X; Liang JJ; Zhao L; Wang Z Multiobjective Optimization of Ethylene Cracking Furnace System Using Self-Adaptive Multiobjective Teaching-Learning-Based Optimization. *Energy* 2018, 148, 469–481.
- (24). Nath R; Alzein Z On-Line Dynamic Optimization of Olefins Plants. *Comput. Chem. Eng.* 2000, 24, 533–538.
- (25). Thiruvenkataswamy P; Eljack FT; Roy N; Mannan MS; El-Halwagi MM Journal of Loss Prevention in the Process Industries Safety and Techno-Economic Analysis of Ethylene Technologies. *J. Loss Prev. Process Ind.* 2016, 39, 74–84.
- (26). Zhao H; Ierapetritou MG; Rong G Production Planning Optimization of an Ethylene Plant Considering Process Operation and Energy Utilization. *Comput. Chem. Eng.* 2016, 87, 1–12.
- (27). Chen M; Xu Q Upset-Conscious Scheduling for Continuous Parallel-Process and Performance Decaying Unit System. *Chem. Eng. Sci.* 2019, 195, 828–840.
- (28). Wauters S; Marin GB Kinetic Modeling of Coke Formation during Steam Cracking. *Ind. Eng. Chem. Res.* 2002, 41 (10), 2379–2391.
- (29). Bi K; Qiu T Novel Naphtha Molecular Reconstruction Process Using a Self- Adaptive Cloud Model and Hybrid Genetic Algorithm – Particle Swarm Optimization Algorithm. *Ind. Eng. Chem. Res.* 2019, 58, 16753–16760.
- (30). Wernicke S Efficient detection of network motifs. *IEEE ACM T. Comput. Bi.* 2006, 3(4), 347–359.
- (31). Wernicke S; Rasche F FANMOD: a tool for fast network motif detection. *Bioinformatics* 2006, 22(9), 1152–1153. [PubMed: 16455747]
- (32). Szegedy C; Liu W; Jia Y; Sermanet P; Reed S; Anguelov D; Erhan D; Vanhoucke V; Rabinovich A Going Deeper With Convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2015; pp 1–9.
- (33). Rossi F; Rovaglio M; Manenti F Chapter 18 - Model Predictive Control and Dynamic Real-Time Optimization of Steam Cracking Units In *Mathematical Modelling of Gas-Phase Complex Reaction Systems: Pyrolysis and Combustion*; Faravelli T, Manenti F, Ranzi E. Eds.; Elsevier, 2019; Vol. 45, pp 873–897.
- (34). Heynderickx GJ; Cornelis GG; Froment GF Circumferential Tube Skin Temperature Profiles in Thermal Cracking Coils. *AIChE J.* 1992, 38 (12), 1905–1912.
- (35). Fakhroleslam M; Sadrameli SM Thermal cracking of hydrocarbons for the production of light olefins; A review on optimal process design, operation, and control. *Ind. Eng. Chem. Res.* 2020, just accepted.
- (36). Li C; Zhu Q; Geng Z Multi-objective particle swarm optimization hybrid algorithm: an application on industrial cracking furnace. *Ind. Eng. Chem. Res.* 2007, 46(11), 3602–3609.
- (37). Salari D; Niaei A; Nabavi R Multi-objective genetic optimization of ethane thermal cracking reactor. *Iran. J. Chem. Eng.* 2008, 5, 29–39.
- (38). Nabavi S; Rangaiah G; Niaei A; Salari D Multiobjective optimization of an industrial LPG thermal cracker using a first principles model. *Ind. Eng. Chem. Res.* 2009, 48, 9523–9533.

- (39). Fang Z; Qiu T; Zhou W Coupled Simulation of Recirculation Zonal Firebox Model and Detailed Kinetic Reactor Model in an Industrial Ethylene Cracking Furnace. *Chinese J. Chem. Eng.* 2017, 25 (8), 1091–1100.
- (40). Masih M; Algahtani I; Mello L De. Price Dynamics of Crude Oil and the Regional Ethylene Markets. *Energy Econ.* 2010, 32 (6), 1435–1444.
- (41). Bi K; Zhang C; Qiu T An Ingenious Characterization of Reaction Network Using Sub-Network Reconstruction. *Comput. Chem. Eng.* 2020, 134, 106695.
- (42). Abadi M; Barham P; Chen J; Chen Z; Davis A; Dean J; Devin M; Ghemawat S; Irving G; Isard M; Kudlur M; Levenberg J; Monga R; Moore S; Murray DG; Steiner B; Tucker P; Vasudevan V; Warden P; Wicke M; Yu Y; Zheng X TensorFlow: A System for Large-Scale Machine Learning In 12th Symposium on Operating Systems Design and Implementation; Association: Savannah, GA, 2016; pp 265–283.
- (43). Bordes A; Bottou L; Gallinari P SGD-QN: Careful Quasi-Newton Stochastic Gradient Descent. *J. Mach. Learn. Res.* 2009, 10 (59), 1737–1754.

Figure 1.

The integrated framework for transfer learning and intelligent heuristic optimization.

Figure 2.
A three-parameter COT curve during production.

Figure 3.
The neural network structure of steam cracking prediction.

Figure 4.
The transfer learning process of steam cracking neural network

Figure 5.
Schematic diagram of the millisecond cracking furnace

Figure 6.
Optimized COT curve for different sets of pricing weights.

Figure 7.
Product distribution for different sets of pricing weight.

Figure 8.
 $OBJ - x_1 - x_2$ curve for the different sets of pricing weights (a) Set 1; (b) Set 2; (c) Set 3.

Figure 9.
Node information of propylene/butadiene from reaction network characterization.

Figure 10.

The reaction routes diagram for steam cracking with the increasing COT.

Table 1.

Detailed setting of the GA-PSO algorithm.

PSO settings (50 individuals)	velocity	$v_{ij}(t+1) = \omega(t) \times v_{ij}(t) + c_1 \times rand \times (p_{bj}(t) - x_{ij}(t)) + c_2 \times rand \times (g_{bj}(t) - x_{ij}(t))$				
	position	$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1)$				
	ω	$\omega = \frac{\omega_{max} - \omega_{min}}{1 + \exp(2k_w t / t_{max} - 6)} + \omega_{min}$				
	c1	$c_1(t) = c_{1max} - \frac{(c_{1max} - c_{1min})}{k} \times (\arctan(\frac{20t}{t_{max}} - \theta_1) + \arctan\theta_1)$				
	c2	$c_2(t) = c_{2max} - \frac{(c_{2max} - c_{2min})}{k} \times (\arctan(\frac{20t}{t_{max}} - \theta_1) + \arctan\theta_2)$				
	k	$\arctan(20 - \theta_1) + \arctan\theta_2$				
	ω_{max}	0.9	c_{1max}	2.4	c_{2max}	2
	ω_{min}	0.4	c_{1min}	1	c_{2min}	0.5
	k_w	10	θ_1	4	θ_2	6
GA settings (50 individuals)	Pc	$\frac{P_c^1(f_c - f_{avg}) + P_c^2(f_{max} - f_c)}{f_{max} - f_{avg}} f_c > f_{avg}$				
		$\frac{P_c^2(f_c - f_{min}) + P_c^3(f_{avg} - f_c)}{f_{avg} - f_{min}} f_c \leq f_{avg}$				
	Pm	$\frac{P_m^1(f_m - f_{avg}) + P_m^2(f_{max} - f_m)}{f_{max} - f_{avg}} f_m > f_{avg}$				
		$\frac{P_m^2(f_m - f_{min}) + P_m^3(f_{avg} - f_m)}{f_{avg} - f_{min}} f_m \leq f_{avg}$				
	P_c^1	0.4	P_c^2	0.7	P_c^3	1
	P_m^1	0.002	P_m^2	0.004	P_m^3	0.006
	Ggap	0.9	coding	binary	crossover	one-point

Table 2.

Weight sets for main products in steam cracking.

Product	C ₂ H ₄	C ₃ H ₆	C ₄ H ₆	C ₆ H ₆	C ₇ H ₈	C ₈ H ₁₀	CH ₄	H ₂	C ₂ H ₆	Residue
Weight set 1	1.00	0.84	0.84	0.78	0.76	0.80	0.16	1.30	0.36	0.42
Weight set 2	1.00	1.20	1.16	0.97	0.94	0.99	0.20	1.61	0.44	0.53
Weight set 3	1.00	1.04	1.04	1.11	1.08	1.16	0.20	1.61	0.44	0.53

Table 3.

Profit margin improvement after the optimization compared to base case.

	Set 1	Set 2	Set 3
Base case	0.6742	0.8093	0.7770
Optimized case	0.6864	0.8176	0.7882
Profit margin improvement	1.81%	1.03%	1.44%

Table 4.

Reaction rate fluctuation trend with the increasing COT for example reactions.

Index	Reaction	Trend
1	$C_2H_5\bullet + C_4H_6 \rightarrow C_2H_4 + SC_4H_7\bullet$	Strongly accelerated
2	$C_2H_5\bullet + C_3H_6 \rightarrow 1.5 C_2H_4 + 0.5 NC_3H_7\bullet + 0.5 CH_3\bullet$	Strongly accelerated
3	$C_4H_5\bullet + NC_6H_{14} \rightarrow C_4H_6 + C_3H_6 + NC_3H_7\bullet$	Slightly accelerated
4	$C_4H_5\bullet + NC_6H_{14} \rightarrow C_4H_6 + NC_5H_{10} + CH_3\bullet$	Slightly accelerated
5	$H\bullet + CH_3\bullet \rightarrow CH_4$	Slightly hindered
6	$H\bullet + NC_6H_{14} \rightarrow NC_3H_7\bullet + C_3H_6 + H_2$	Slightly accelerated
7	$H\bullet + NC_6H_{14} \rightarrow CH_3\bullet + NC_5H_{10} + H_2$	Slightly accelerated
8	$NC_6H_{14} \rightarrow CH_3\bullet + NC_3H_7\bullet + C_2H_4$	Slightly accelerated

Table 5.

Detailed training performance of the transfer learning method.

Configuration		Neural network pre-training for naphtha 1 in furnace 1			Transfer tasks for naphtha 2 in furnace 2		
CPU	GPU	Speed	Training loss	Testing loss	Speed	Training loss	Testing loss
i7-4770 @3.40GHz	/	~25h	3.36	3.62	~4.2h	4.06	4.33
i7-9750 @4.50GHz	RTX 2070mq (8GB)	~2.6h	3.27	3.40	~28m	3.97	4.21
Xeon Phi 5110P @2.50GHz	Tesla V100 (32GB)	~1.2h	3.29	3.39	~17m	3.92	4.17

Table 6.

A prediction comparison of the transfer learning model and the EcSOS.

Product	Set 1			Set 2			Set 3		
	Predicted	Simulated	Error	Predicted	Simulated	Error	Predicted	Simulated	Error
H ₂	1.2580	1.2021	4.65%	1.1065	1.1454	−3.40%	1.0974	1.1423	−3.93%
CH ₄	16.8642	17.1628	−1.74%	15.7530	16.2101	−2.82%	15.1267	15.8146	−4.35%
C ₂ H ₄	31.2485	30.7958	1.47%	30.0416	30.1562	−0.38%	29.7305	29.9250	−0.65%
C ₄ H ₆	3.9218	4.1247	−4.92%	4.4317	4.2777	3.60%	4.3438	4.1759	4.02%
C ₃ H ₆	16.5323	16.8284	−1.76%	17.3762	16.9656	2.42%	17.3850	17.0257	2.11%
C ₂ H ₆	1.6760	1.6924	−0.97%	1.8020	1.7390	3.62%	1.8683	1.8051	3.50%
C ₆ H ₆	3.7971	3.9744	−4.46%	4.3207	4.1553	3.98%	4.5231	4.3884	3.07%
C ₇ H ₈	3.0051	2.8931	3.87%	3.1194	3.0564	2.06%	3.2937	3.1452	4.72%
C ₈ H ₁₀	2.6210	2.7035	−3.05%	2.6770	2.8126	−4.82%	2.7766	2.8878	−3.85%

Table 7.

Detailed performance of the GA-PSO optimization process.

Set	1	2	3
Average time-consumption (s)	108.6	38.0	86.2
Probability of finding optimal solution	90%	100%	100%
Average function calls	4390	1420	3270