

実施計画書

東京大学・大学院工学系研究科化学システム工学専攻 金子弘昌

提案テーマ名：

石油精製工場におけるビッグデータを活用した安全かつ効率的なプラント管理手法の確立

■「実施内容」および「スケジュール」

○ 平成28年度：ビッグデータを扱う情報基盤およびソフトセンサーの開発

・ビッグデータ収集（1Q、2Q）

比重・蒸留温度などの原油の成分分析結果、毎分の温度・圧力などのセンサーデータ、最終製品の性状データなどの石油精製工場のデータを収集する。テキストデータがあればそれをデジタル化して時系列データと同様に整理する。複数の企業のデータを収集するだけでなく、論文や各種データベースからも収集し、それらを統合する。

・ビッグデータ情報基盤構築（2Q、3Q）

外れ値検出、欠損値の補間、時系列データの平滑化などのデータの前処理技術を開発する。

外れ値はソフトセンサーの予測精度を低下させる可能性があるため、適切に外れ値を検出する手法を開発する。まず収集したデータをモデル構築用データとモデル検証用データに分割し、モデル検証データの予測精度が向上するように、モデル構築用データから外れ値を除去する。外れ値の検出には、時系列データが時間的に前後のデータ間で類似しているという特徴を活用し、前後データを使用した平滑化結果との差分を計算して外れ値を検出する。これにより外れ値検出のみならず時系列データの平滑化も同時に達成される。モデル検証用データの予測精度をモニタリングしながら外れ値を除去することで、情報量の大きい必要なデータが外れ値となることを防ぐ。もちろんその知見を用いて最終的にはモデル検証用データの外れ値も検出する。検出された外れ値は欠損値として処理する。

センサーデータのように高頻度で測定されている変数や、成分データ・性状データのように測定頻度の低い変数が存在する。つまりデータセットが虫食い状態といえる。値が揃っているデータのみを使用することも考えられるが、出力変数と入力変数との間の相関関係をモデル化するにはより多くのデータを使用できるほうが望ましい。そこで変数間の関係および時系列データの特徴をふまえて、機械学習および EM アルゴリズムを活用してモデル構築と同時にデータの補間を行う手法を開発する。このときにモデルの予測性能が向上するようにデータ補間とモデル構築を行うべきである。そこでモデルの予測性能を評価する指標が必要となる。一般的にモデルの予測性能を評価する際にクロスバリデーションが行われる。しかしデータが揃っていないことからクロスバリデーションでは適切にモデルの性能を評価できないと考えられる。その際はデータの midpoint に基づく指標を使用したり、新たな予測性能の指標を開発することで対応する。

・ソフトセンサー構築（1Q、2Q）

センサーデータおよび運転条件から最終製品の性状を推定するソフトセンサーを構築する。プロセス変数間に非線形性があり高精度なモデルが構築できない場合はプロセスモデルを組合せたり非線形の回帰分析手法を駆使して非線形性に対応する。運転データのすべてのデータ範囲で精度良く推定できるモデルを構築することは困難であり、データの粗密によってモデルの推定精度は異なる。そこで、モデル構築用データのデータ範囲からの差を表現することで、推定値にエラーバーを付加する。これにより確率分布としてラボ分析データを推定する。この確率分布は後述するように異常検出モデルにおいて異常が起きる確率を表現できる。

ソフトセンサーの予測精度向上のためには、プロセスの動特性を考慮しなければならない。そこでデータセットに時間遅れのプロセス変数を追加することで動特性を考慮する。ただ、出力変数と関係のない余計な変数はノイズとなり逆に予測精度を低下させる。そこで変数間の相関関係やモデル検証用データの予測性能を考慮して、必要な変数のみ選択する手法を開発する。なお、データ数および変数の数が膨大であり、一般的な回帰分析手法で対応できない場合は、ディープラーニングを駆使してソフトセンサーモデルを構築する。さらに、ソフトセンサーモデルの性能の劣化を低減するため、新しい測定データを活用してモデルの自動メンテナンスを行う。つまりオンライン学習するディープラーニングの新規手法を開発する。

○ 平成29年度：将来の異常予測および異常診断技術の開発

・異常検出モデル構築

新しい運転データが得られた際に、プラントの状態を診断して異常を検出するモデルの構築手法を開発する。柔軟な異常検出を達成するため、正常・異常の判定を行うのではなく異常が起こる確率で表現する。前年度に構築したソフトセンサーでリアルタイムに推定でき、その推定結果を確率分布で出力することができる。そのため、管理する出力変数に上限・下限を設け、確率分布を上限から無限大および下限から負の無限大まで積分することで、対象の変数が上限・下限を越える確率を計算できる。実際には一変数ごとに管理するわけではなく、複数の変数を同時に管理するため多変量プロセス管理手法を用いる。

線形の主成分分析で対応できない場合はカーネル法や one-class support vector machine を活用して非線形性に対応する。いずれの多変量プロセス管理手法を用いる場合でも、ある管理指標に変換することになるため、上述したように異常が起きる確率を求めることができる。異常検出の精度が低い場合はアンサンブル学習と合わせて精度の向上を試みる。

また本研究では、異常検出の精度向上および後の異常原因の診断のため、既存の異常データを活用する。正常データと異常データとの間の領域における異常検出性能の低下を避けるため、正常データや異常データの種類ごとに多変量プロセス管理手法を用いる。これにより、異常の検出ができるだけでなく、どの異常か判定することも可能となる。もちろん正常状態でも既存のどの異常状態でもない場合は未知の異常として処理され、後に異常原因の診断が行われる。

本研究ではデータが測定された時刻の異常検出だけでなく、将来起こりうる異常の検出も実施する。将来のプロセス状態を推定するためプロセスのダイナミクスを考慮する。これまでのデータの軌跡からデータが変化する速度を計算する。データの位置と速度より、正常状態から異常状態に向かっているのか否か、およびあとどれくらいで異常状態に到達するかを推定できる。現在のデータから前の時刻のデータの差分を取ることで単純にデータの変化を計算できるが、必要に応じてさらに前の時刻からの変化や、プロセスモデルを用いてプロセスのダイナミクスを推定するモデルを構築する。

プラントの状態は時々刻々と変化しているため、新しい測定データを用いてモデルを更新するシステムも開発する。

・異常診断モデル構築

異常検出モデルで将来の異常が検出された後に、その異常原因を診断する方法を開発する。既存の異常データと同様の異常であれば、異常検出モデルの出力から診断可能である。異常検出モデルにおいて未知の異常と診断された場合は、まず各プロセス変数の異常への寄与率を求める。異常検出モデルは非線形のプロセスモデルが一部に使用されていたり非線形の機械学習手法を使用したりするため従来の方法では寄与率を計算できない。本研究ではモデルを微分して該当するデータにおける微分係数を求めることで、各プロセス変数の異常状態での変化の割合を計算する。この値の絶対値が大きいプロセス変数ほど異常に寄与していると考えられる。

実際のデータにおいて各プロセス変数の異常への寄与の大きさを求めた後に、プロセスモデルやソフトセンサーで構築される擬似プラントモデルで様々な異常原因を模倣して実データと同様にして各プロセス変数の異常への寄与の大きさを計算する。実データでの各変数の寄与の大きさのパターンと擬似プラントモデルでのそれを比較し、それらが類似した場合には擬似プラントモデルでの異常原因が実際の異常原因と一致すると考えられる。

異常原因が判明した後に同様の異常を擬似プラントモデルで発生させる。その後、異常状態から正常状態への操作方法を、擬似プラントモデルを逆に解析することで検討する。まず操作変数の操作方法の候補をコンピュータ上で仮想的に多数生成し、擬似プラントモデルへ入力して操作中の擬似プラントの変化および操作終了後のプラント状態を確認する。候補の中から擬似プラントが安定的に正常状態に推移した操作方法の候補を選択する。この方法を用いることで、擬似プラントのシミュレーションにより実プラントでの安定的な正常状態への制御方法を検討できる。プラントにおける設定値変更のみならず外乱がある中での安定運転にも対応する。

操作方法の候補の生成の仕方にプロセスの知見を含めることで、より現実的な操作方法を検討できる。操作方法の候補が非常に多くすべてを検討できない場合は遺伝的アルゴリズムなどの最適化手法で対応する。

○ 平成30年度：ペトロレオミクス技術を活用したリアルタイム最適化手法の開発

・ベイズ推定技術の開発

ビッグデータから構築される、製品品質を推定するソフトセンサーモデルおよび原油特性を推定するソフトセンサーモデルを用いることで、原油ごとの原油の成分分析結果・各化合物の成分組成比に応じて、目標とする最終製品品質および製造コストを達成できる運転条件を探索する手法を開発する。なお原料特性推定モデルを構築する際には、ペトロレオミクス技術および各化合物の物性データを活用する。つまり、原料特性推定モデル構築の際に化学構造データと物性データとの相関関係を、機械学習手法を駆使してモデル化することで、原料特性推定モデルの精度を向上させる。

X を運転条件、Z を原油特性、Y を最終製品品質とする。データベースから求められる運転条件の事前分布 $p(X)$ 、原油特性の事前分布 $p(Z)$ 、原油特性推定モデル $p(Z|X)$ 、製品品質推定モデル $p(Y|X,Z)$ および $p(Y|Z)$ から、ベイズの定理により、製品品質および原油特性が与えられたときに運転条件の事後分布を計算する（図 1 参照）。製品品質については目標値を入力し、原油特性は使用している原油のパラメータの値を入力することで、そのために必要な運転条件の候補を出力できる。

線形のソフトセンサーモデルであれば $p(X|Y)$ の定式化が可能であるが、非線形の物性推算モデルにおいて $p(Y|X,Z) \cdot p(Y|Z) \cdot p(Z|X)$ を正規分布もしくは混合正規分布で表現できないことが問題として考えられる。その場合、一つの実線形モデルではなく複数の線形モデルを組み合わせて解決する。これにより非線形性を表現可能であり、各線形モデルは $p(Y|X,Z)$ を正規分布で表現できる。もう一つの対策としては、モンテカルロ・シミュレーションにより事後分布を計算する。

X: 運転条件, Z: 原油特性, Y: 製品品質

$$p(X|Y,Z) = \frac{p(Y|X,Z)}{p(Y|Z)} \frac{p(Z|X)}{p(Z)} p(X)$$

製品品質 原油特性
推定モデル 推定モデル

図 1. リアルタイム運転条件最適化のためのベイズ推定

・リアルタイム最適化手法開発

上で計算したデータベースから $p(Y|X,Z)$ に基づいて、目標とする最終製品品質および製造コストを達成できる運転条件をリアルタイムに最適化する手法を開発する。ソフトセンサーモデルが非線形であったり設計変数が多い場合にモンテカルロ・シミュレーションを使用する際には計算コストがかかるため、リアルタイムに最適化できるよう効率的な最適化法を開発する。これにより石油精製工場におけるリアルタイム最適化が達成される。

■「支出計画」

・設備費 計 1300 千円

高性能コンピュータ 1000 千円

外付けハードディスク 300 千円

・物品費 計 1300 千円

ノートパソコン 200 千円

持ち運び用ハードディスク 100 千円

ソフトウェア 500 千円

データベース 500 千円

・消耗品費 計 750 千円

論文投稿料 600 千円

図書 150 千円

・旅費 計 2400 千円

国内発表・打合せ6回 600 千円

国外発表・打合せ6回 1800 千円

・外注費 計 300 千円

データ入力の外注 300 千円

・諸経費 計 150 千円

一年あたり 50 千円 150 千円