

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 26 日現在

機関番号：17102

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26420781

研究課題名(和文) 正常時運転履歴データを活用した化学プラントの状態監視およびソフトセンサーへの応用

研究課題名(英文) State Monitoring of Chemical Plants using Normal Operation Historical Data and Application to Soft Sensors

研究代表者

柘植 義文 (Tsuge, Yoshifumi)

九州大学・工学研究院・教授

研究者番号：00179988

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：化学プラントの運転・監視のためにリアルタイムで状態変数の値を測定すると同時に運転履歴データとして保存されているが、そのほとんどは正常運転時の履歴データである。本研究では、大量に蓄積されている過去の正常運転時の履歴データを参照して、現在のあるべき正常状態を的確に予測する手法を開発し、ボイラープラントのように正常状態が不確定に変動するプラントに適用できることを示した。さらに、リアルタイムでの測定が困難な変数の値を推定するために広く用いられているソフトセンサーへ応用できることを実プラントでの運転実績データに基づいた数値実験により示した。

研究成果の概要(英文)：In a chemical plant, the values of the state variables are measured in real time for operation and monitoring and are stored as the operation historical data at the same time. But most of them are the historical data during normal operation. In this study, we have developed a method to accurately predict the current normal state by referring to historical data during past normal operation which has been accumulated in large quantities. Its usefulness has been demonstrated by applying it to the boiler plant whose normal states are uncertainly varied. Furthermore, we showed that the proposed method can be applied to soft sensors widely used for estimating values of variables that are difficult to be measured in real time by numerical experiments based on operational data in actual plant.

研究分野：プロセスシステム工学

キーワード：化学プラント 運転監視 正常状態予測 ソフトセンサー データベースモデル

## 1. 研究開始当初の背景

化学プラントの運転・監視に分散型制御システム(Distributed Control System: DCS)を利用するのが一般的であり、安全・安心で高効率な運転を実現するために、高度制御システムやアラームマネジメントシステムなどが導入されている。同時に、監視や制御のために測定された状態変数の値は運転履歴データとして、蓄積・保存されている。このようなデータの利用例として、運転中に発生したトラブルを収束させた後で、トラブルの原因解析のためにトラブル時の運転履歴データを検証に利用することはある。しかし、保存されている運転履歴データの大部分は正常運転時のデータであり、順調な生産が継続されている間は単純に蓄積・保存されているだけのことも多いが、正常運転時の履歴データが全く利用されていない訳でもない。例えば、近年利用されているソフトセンサーでは、目的変数と説明変数との間の関係を表現するモデルを構築する必要があり、その学習データとして正常運転時の履歴データの一部を利用している。そのため、学習データとは異なる運転状態でのソフトセンサーの出力値(目的変数の値の推定値)の誤差が大きくなるという問題点があり、学習データの選定やモデル更新のタイミングなどの課題が残っている。あるいは、定常運転を行う連続プラントでの異常値の検出法としてよく用いられる3法での標準偏差の決定にも正常運転時の履歴データの一部を利用している。この場合、正常運転時だからと言って常に同じような値を取るとは限らず、長期間の連続運転中には季節変動や生産条件の変更等などのために運転条件は多少なりとも変動して正常状態も揺らいでいるのが現実であるが、正常状態の揺らぎは標準偏差に丸め込まれている。結局のところ、正常運転時の履歴データが対象プラントの動的な特性に従って得られているにも関わらず、動的な特性に関する情報は有効利用されていない。さらに、ボイラープラントのようなユーティリティプラントでは、正常運転時でもスチームの必要供給量が変動するために正常状態が一定でなく、3法的な異常検知は行えない。以上のように、正常運転時の履歴データが大量に蓄積・保存されているにも関わらず、それを有効に活用しているとは言い難い。

近年、他分野においてビッグデータから有益な情報を抽出して販売や生産活動に活用する事例が報告されているが、用いられている手法はデータマイニングの手法である。以前、著者らはデータマイニングの手法を正常運転時の履歴データに適用して、正常運転時のオペレータのノウハウを抽出し正常運転時の状態予測や監視等に利用する方法を検討したが、十分な成果は得られなかった。その理由は、履歴データはプラントの動特性に従って推移した状態を時間順に並べた時系

列データであり、その順番に意味があるのに対し、データマイニングの手法ではその順番を考慮することができないためであった。

## 2. 研究の目的

本研究では、履歴データが時系列データであることを考慮して正常運転時の履歴データを最大限活用する新たな正常状態予測手法を提案し、大規模化学プラントの運転監視、異常検知、異常診断に役立てる方法を開発することを目的とする。特に、正常状態予測法の精度の事前評価法とデータベース(DB)の更新法について検討する。また、近年、その有用性が認められ、多くの工場で導入されているソフトセンサーへの応用も試み、従来用いられている手法との比較検討を行い、本手法の有用性を検討することを目的とする。

## 3. 研究の方法

## (1) 正常状態予測法の提案

正常運転時の履歴データを有効利用する方法についての定式化は済んでいる。すなわち、 $L$ 組の正常運転時の履歴データをデータベース(DB)に格納する。 $p$ 番目の履歴データの時刻 $k$ における $r$ 次元入力変数 $U(p, k)$ と $m$ 次元出力変数ベクトル $Y(p, k)$ を次式で表す。

$$\begin{aligned} U(p, k) &= [u_1(p, k), u_2(p, k), \dots, u_r(p, k)] \\ Y(p, k) &= [y_1(p, k), y_2(p, k), \dots, y_m(p, k)] \\ & \quad (p = 1, 2, \dots, L; k = 0, 1, 2, \dots, K(p)) \end{aligned} \quad \dots (1)$$

格納されている正常状態を次式で定義し、レコードと呼ぶ。

$$Z(p, k) = [Y(p, k-1), U(p, k)] \quad \dots (2)$$

$$(p = 1, 2, \dots, L; k = 1, 2, \dots, K(p) - 1)$$

ある時刻 $t$ 以降の入力ベクトル $U(t)$ と出力ベクトルの予測値 $\hat{Y}(t-1)$ が与えられとき、それらをモデル入力 $S(t) = [U(t) \ \hat{Y}(t-1)]$ とし、これと類似したレコードを探索する。探索結果は類似度の定義に依存するが、一般的にはベクトル空間内のユークリッド距離やマハラビノス距離を基準にし、参照距離 $d_{ref}$ 以下を類似レコードとする。その後、所定の方法により、類似レコードの出力変数の値からモデル出力 $Y(t)$ (時刻 $t$ の予測値)を算出する。以上のように、原理は非常に単純であり、高次モデルへの拡張も容易に行える。従って、実際にシステム化するときは、モデル次数、類似度の定義、およびモデル出力の算出法が鍵となる。それらを踏まえて具体的なシステム化に取り組んだ。

システムの検証には、ダイナミックシミュレータで作成した正常運転時の履歴データを利用することにして、負荷変動があり正常状態が大きく変わるボイラープラントを対象にした。まず、ボイラープラントから供給される3種類の蒸気(高圧・中圧・低圧)の蒸気を変動する運転履歴データを350日分作

成した（測定変数は 99 個で測定間隔は 1 分である．これを履歴データ X と呼ぶ）．一方，検証用データとして次の 3 種類のデータを作成した．

DataA: 履歴データに格納されているデータと同様な運転状態の検証データ．過去に経験した変動量の負荷変動が含まれている．

DataB1,B2: 履歴データに格納されているデータと異なる運転状態の検証データ．過去に経験していない変動量の負荷変動が含まれている．B1 と B2 に関しては互いに同じ変動量の負荷変動を生じさせているが，変動のパターンが異なる．

DataC: DataB と同様に，履歴データに格納されているデータと異なる運転状態の検証データ．DataB1 と同じ負荷変動量の負荷変動を与えている．

実用化に向けての課題であった正常状態の予測精度の事前評価法として，leave-one-out cross-validation (LOOCV)を採用し，その際に得られる各出力変数の絶対予測誤差の最大値  $e_{j,max}$  を，予測精度の指標とした．DB 内には過去の正常運転時の履歴データを格納しているので，これまでと同様の運転状態であれば LOOCV により得られた最大予測誤差の範囲で予測が可能と考えられる．一方，プラントの運転状態が変化して過去の運転と異なる状態になると予測精度が低下する可能性がある．そのため予測精度を維持するためには，運転状態の変化に合わせた DB の更新が求められる．本研究では，LOOCV により得られた最大予測誤差よりも大きな誤差が生じた際に，正常な状態であれば運転状態が変化したとみなし，DB の更新を行う．

#### (2) ソフトセンサーへの応用

ソフトセンサーの目的変数はリアルタイムで測定されていなく間欠的な分析等で測定値が得られる出力変数であり，説明変数はリアルタイムで測定されている入出力変数から選択された変数である．従って，レコードの定義を変更することによって比較的簡単に正常状態予測用のモデルの考え方をソフトセンサー用に応用することができる．なお，一般的に利用されているソフトセンサーは目的変数毎にモデル式が構築されているが，本研究での手法を利用する場合は，複数の目的変数を同時に処理することも可能である．その点を踏まえてソフトセンサー用にシステムを改良した．

検証には，某石油精製会社から頂いた 4 年間の運転実績データ（2007 年 7 月 1 日 8 時～2011 年 2 月 28 日 7 時までの 1 時間平均データ 27630 サンプルであり，測定値は運転データ内の最大値と最小値で [0 1] に規格化されている）を対象にした．提供された運転停止/開始等のイベント情報より運転停止期間とスタートアップ期間を除外し，正常運転

期間を抽出した．さらに変数が運転停止直前やスタートアップ後に 0 をとり続けている期間も運転停止期間と同様に除外し，残りの 26340 サンプルを図 1 のように区分した．図 1 中の塗りつぶされていない期間が正常運転と判断した期間(Normal1～Normal3)である．説明変数は 107 変数，目的変数は  $y_1 \sim y_7$  の 7 変数であり，移動平均に基づいた外れ値処理手法を正常運転期間に適用した．比較対象のモデルとしては，ソフトセンサーとして広く用いられている PLS 法(変数選択法として VIP と BETA の 2 種類)と Lasso 法を採用した．

提案手法をソフトセンサーに応用するにあたり，全ての説明変数を利用して 7 つの目的変数の値を個別に推定することも同時に推定することも可能である．また，一般的な多変量解析に基づいたソフトセンサーと同様に説明変数の選択を行うことも可能である．それらが推定精度に及ぼす影響についても検討した．

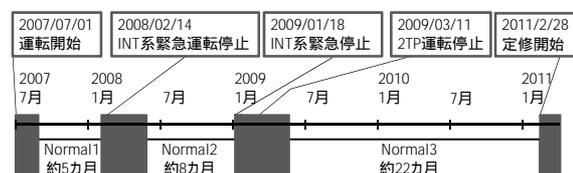


図 1 運転データの分類

#### 4. 研究成果

##### (1) 正常状態予測法の検討結果

ボイラープラントシミュレータでの検討結果を以下に取り纏める．

まず，最初に準備した履歴データ X での LOOCV の最大予測誤差に対して，検証データ DataA での予測誤差は全ての変数において LOOCV の最大予測誤差以下の範囲に収まった．すなわち，DataA の様な過去の正常運転履歴データと同様な運転状態の予測では，LOOCV によって得られた値を想定される予測誤差の最大値として扱うことができると考えられる．

一方，検証データ DataB1 および DataB2 においては，LOOCV の最大予測誤差を超えた変数が多数存在した．そこで，履歴データ X に DataB1 のみを追加したものを履歴データ Y，DataB1 と B2 の両方を追加したものを履歴データ Z として，DB の更新の影響を検討した．表 1 に履歴データ X，Y，Z での最大予測誤差の結果を示す．更新前の履歴データ X に比べて，履歴データ Y での LOOCV の値は精度を低く評価する変数の数が増えたが，履歴データ Z では履歴データ X と同程度に小さくなった．

検証データ DataC については，履歴データ X を用いた場合には LOOCV での値を超える変数も見受けられたが，履歴データ Y や Z を用いた場合には LOOCV での値以下の範囲で予測できた．一例として，脱気器給水流量

FC1101 の予測結果を図 2 に示す．390min 付近において、履歴データ X では予測値の追従できていないのに対して、履歴データ Y, Z では予測値の追従ができており、予測誤差が小さくなっていることがわかる．DB を更新することで、DB モデルの精度が向上している．

表 1 LOOCV での最大予測誤差

最大予測誤差	変数の数		
	履歴データ X	履歴データ Y	履歴データ Z
0.20~0.30	0	3	0
0.10~0.20	5	6	7
0.05~0.10	11	13	10
0.01~0.05	29	38	32
0.00~0.01	54	39	50

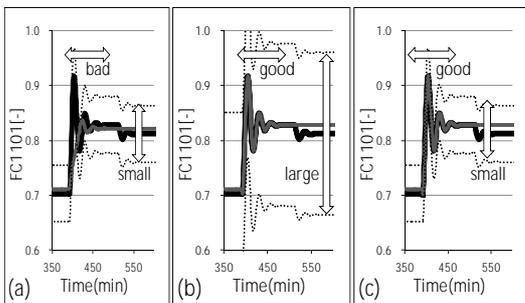


図 2 DataC における FC1101[-]の予測結果  
(a)履歴 X (b)履歴 Y (c)履歴 Z

(2) ソフトセンサーへの応用の検討結果

推定性能を比較するため、次の 3 通りの推定を行った．

- (i) Normal1 の前半 3 ヶ月を学習データ、残り 2 ヶ月を検証データ
- (ii) Normal2 の前半 3 ヶ月を学習データ、残り 5 ヶ月を検証データ
- (iii) Normal3 の前半 11 ヶ月を学習データ、後半 11 ヶ月を検証データ

推定精度の指標として Root Mean Square Error (RMSE)を用い、図 3(i) ~ (iii)に結果を示す．目的変数ごとにおおまかに比較すると、提案手法は比較手法と同程度の RMSE 値を示している、詳細にみると、提案手法の推定精度が全ての比較手法より劣っている場合もあった(図 3(i)の  $y_7$  や図 3(iii)の  $y_2$ )．これは、DB 内に蓄積されていない値が検証データに含まれていたためである．逆に、比較手法の推定精度がより悪くなる場合があった．これは、外れ値や学習データ選定の影響であった．

学習データの選定による推定結果の違いを検証するため、最も期間の長い Normal3 を用いた(iii)中の学習データをさらに区分し、(iv) Normal3 前半初期 5 ヶ月を学習データ、(v) Normal3 前半中間 5 ヶ月を学習データ、(vi) Normal3 前半後期 5 ヶ月を学習データ、で、後半 11 ヶ月を検証データとする 3 通りの推定を行った．その結果、比較手法では学習データのとり方によって大きく推定精度が変化する目的変数がいくつかあったのに

対して、提案手法では学習データが変わっても推定精度は大差なかった．

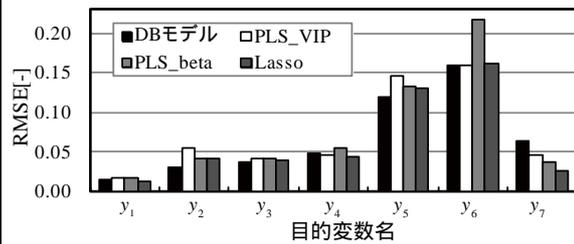


図 3(i) (i)の場合の目的変数の RMSE 値

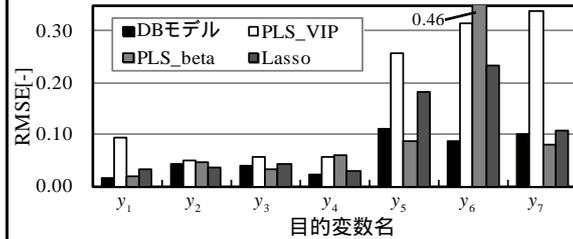


図 3(ii) (ii)の場合の目的変数の RMSE 値

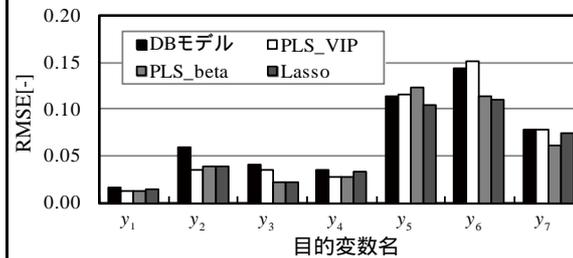


図 3(iii) (iii)の場合の目的変数の RMSE 値

運転停止期間を挟んでも推定が可能かどうか検証するため

- (vii) Normal1 を学習データ、Normal2 を検証データ
- (viii) Normal1~2 を学習データ、Normal3 を検証データ

とする 2 通りの推定を行った．それぞれの推定結果を検証データが同じである(ii)または(iii)と推定精度を比較したところ、提案手法では、運転停止期間を挟んだ場合とそうでない場合で、推定精度は大きく変わらなかった．

最も期間の長い Normal3 の中から定常運転していると思われる期間 5400 サンプルを選定し、前半 2 ヶ月(2000 サンプル)を DB に蓄積する履歴データ、後半 5 ヶ月(3400 サンプル)を推定の検証データとして、提案手法における変数選択法について検討した結果を示す．

まず事前評価として、履歴データに対して分割数 10 のクロスバリデーションによって変数選択基準を算出した．(A)全ての目的変数で共通の変数選択を行うときは MEN (Mean Error Norm) を評価指標とし、(B)目的変数ごとに個別の変数選択を行うときは MAE (Mean Absolute Error) を評価指標とした．また、GA パラメータは、最大世代数 300、染色体数 50、交叉確率 0.90、突然

変異確率 0.005 とした。表 2 に変数選択適用前後の変数選択基準を示す。2 行目には各手法を示しており、All は変数選択前の全説明変数を選択したモデルを指す。共通変数選択基準 MEN を 2~4 列目に、個別変数選択基準 MAE を 5~7 列目に記す。変数選択前後で比較すると、各手法によって変数選択基準を大幅に改善していることがわかる。また、表 3 に各手法によって選択された説明変数の個数を示す。全 107 個の説明変数を、変数減増法では約 15~30 個、GA を利用した手法では約 50~60 個削除している。これらの結果から、変数選択が推定性能の向上、DB に蓄積するデータの削減へ寄与することが示された。ただし、個別変数選択では、各目的変数で選択変数の組み合わせが異なり、各々に応じたモデルが必要となるため、1 つのモデルで推定するよりも計算負荷が増加する。表 4 に履歴データのクロスバリデーションによる推定性能の事前評価結果を示す。事前評価での RMSE は推定性能の期待値を表す。共通変数選択の結果を 3~4 列目に、個別変数選択の結果を 5~6 列目に記す。(A) 共通変数選択に着目して、変数選択前後の推定性能を比較すると、目的変数間に違いが見受けられた。目的変数  $y_1, y_2, y_5, y_6, y_7$  においては両手法とも 10~20% 程度 RMSE 値が減少しており、MEN の減少率と概ね同程度となった。対して、目的変数  $y_3, y_4$  においては RMSE 値がほとんど変化しておらず、MEN の減少率に満たなかった。これは、変数選択基準 MEN に起因すると考えられる。MEN は複数の目的変数の推定誤差をまとめて評価するため、一部の目的変数 ( $y_3, y_4$ ) の影響が緩和され、それらに対して十分な精度向上を期待できない変数選択結果 (モデル) となった。そこで、共通変数選択を適用する際は、モデルの事前評価から各目的変数の推定性能を確認し、適用可能であるかを判断すべきである。なお、今回の数値実験では目的変数  $y_3, y_4$  において、変数選択前よりも RMSE が著しく悪化することはなかったため、適用可能であると判断した。(B) 個別変数選択に着目して、変数選択前後の推定性能を比較すると、全目的変数において、変数減増法で 10~35% 程度、GA を利用した手法で 10~30% 程度 RMSE 値が減少しており、MAE の減少率と概ね同程度となった。一部の目的変数においては、共通変数選択が個別変数選択に劣る結果となったが、変数選択法を適用することによって推定性能向上が期待される。

表 5 に検証データを対象とした推定性能の事後評価結果を示す。なお、表 4 と同様に事後評価での RMSE を記す。(A) 共通変数選択に着目して、変数減増法について、変数選択前後の推定性能を比較すると、目的変数  $y_6, y_7$  においては事前評価で期待される 10~20% 程度 RMSE 値が減少した。目的変数  $y_1, y_3, y_4, y_5$  においては変数選択前と概ね同程度となった。同様に、GA を利用した手法について、変数選択前後の推定性能を比較すると、目的変数  $y_1, y_5, y_6, y_7$  においては事前評価で期待される 10~20% 程度 RMSE 値が減少した。目的変数  $y_3, y_4$  において

は変数選択前と概ね同程度となった。しかし、目的変数  $y_2$  においては、変数減増法で 24%、GA を利用した手法で 10% RMSE 値が増加した。

表 2 変数選択基準

	MEN			MAE		
	All	Stepwise	GA	All	Stepwise	GA
$y_1$	0.0861	0.0743 (-14)	0.0799 (-7)	0.0068	0.0051 (-25)	0.0054 (-21)
$y_2$				0.0330	0.0253 (-23)	0.0289 (-13)
$y_3$				0.0167	0.0129 (-23)	0.0168 (-22)
$y_4$				0.0157	0.0131 (-17)	0.0138 (-12)
$y_5$				0.0323	0.0184 (-43)	0.0216 (-33)
$y_6$				0.0385	0.0298 (-23)	0.0319 (-17)
$y_7$				0.0293	0.0204 (-30)	0.0243 (-17)

表 3 選択変数の個数

	(A) 共通変数選択			(B) 個別変数選択	
	All	Stepwise	GA	Stepwise	GA
$y_1$	107	87	52	92	50
$y_2$				89	48
$y_3$				93	55
$y_4$				88	50
$y_5$				83	50
$y_6$				77	50
$y_7$				78	59

表 4 変数選択法適用結果の事前評価 (RMSE)

	(A) 共通変数選択			(B) 個別変数選択	
	All	Stepwise	GA	Stepwise	GA
$y_1$	0.0086	0.0065 (-25)	0.0069 (-26)	0.0064 (-25)	0.0060 (-31)
$y_2$	0.0441	0.0403 (-9)	0.0427 (-6)	0.0352 (-20)	0.0394 (-11)
$y_3$	0.0212	0.0212 (±0)	0.0211 (-1)	0.0164 (-23)	0.0162 (-23)
$y_4$	0.0221	0.0225 (+2)	0.0235 (+5)	0.0197 (-11)	0.0188 (-15)
$y_5$	0.0461	0.0329 (-29)	0.0311 (-21)	0.0294 (-36)	0.0316 (-31)
$y_6$	0.0502	0.0459 (-9)	0.0502 (-13)	0.0418 (-17)	0.0441 (-12)
$y_7$	0.0355	0.0283 (-20)	0.0355 (-20)	0.0261 (-27)	0.0287 (-19)

表 5 変数選択法適用結果の事後評価 (RMSE)

	(A) 共通変数選択			(B) 個別変数選択	
	All	Stepwise	GA	Stepwise	GA
$y_1$	0.0117	0.0113 (-4)	0.0109 (-7)	0.0079 (-32)	0.0103 (-12)
$y_2$	0.0581	0.0718 (+24)	0.0639 (+10)	0.0704 (+21)	0.0559 (-4)
$y_3$	0.0526	0.0513 (-2)	0.0499 (-5)	0.0543 (+3)	0.0483 (-8)
$y_4$	0.0467	0.0476 (+2)	0.0482 (+3)	0.0473 (+1)	0.0485 (+4)
$y_5$	0.1079	0.1030 (-5)	0.0960 (-11)	0.1051 (-3)	0.0943 (-13)
$y_6$	0.1039	0.0840 (-19)	0.0790 (-24)	0.1029 (-1)	0.0887 (-15)
$y_7$	0.0491	0.0445 (-9)	0.0448 (-9)	0.0422 (-14)	0.0397 (-19)

括弧内の数値は All 基準の変化率[%]

一方, (B)個別変数選択に着目して, 変数減増法について, 変数選択前後の推定性能を比較すると, 目的変数  $y_1$  と  $y_7$  においては事前評価で期待される 10~35%程度 RMSE 値が減少した。また, 目的変数  $y_3, y_4, y_5, y_6$  においては変数選択前と概ね同程度となった。目的変数  $y_2$  においては, 共通変数選択と同様に過剰適合の傾向が見受けられた。同様に, GA を利用した手法について, 変数選択前後の推定性能を比較すると, 目的変数  $y_1, y_3, y_5, y_6, y_7$  においては事前評価で期待される 10~30%程度 RMSE 値が減少した。目的変数  $y_2, y_4$  においては, 変数選択前と概ね同程度となった。

共通変数選択, 個別変数選択ともに, 一部の目的変数においては事前評価で期待される精度向上を確認できなかったが, 各変数選択法を適用することによって, 推定性能が同等または向上する結果が得られた。

## 5. 主な発表論文等

〔学会発表〕(計 4 件)

Yuki WAKASUGI, Naoki KIMURA and Yoshifumi TSUGE, "Monitoring System for Performance Deterioration of Heat Exchanger in Chemical Plant", The 29th International Symposium on Chemical Engineering (ISChE 2016), Miyazaki, Japan, 3 December, 2016

若杉侑生, 木村直樹, 柘植義文, "化学プラント用熱交換器の状態監視", 第59回自動制御連合講演会, 2016年11月11日, 北九州市国際会議場(北九州市)

Kohei TAKAMATSU, Naoki KIMURA and Yoshifumi TSUGE, "Discussion on Similarity Index and Variable Selection in Soft Sensor using the DB Model", The 27th International Symposium on Chemical Engineering (ISChE 2014), Kuala Lumpur, Malaysia, 6 December, 2014

Masahiro INOUE, Naoki KIMURA and Yoshifumi TSUGE, "Comparison of regression models for soft sensor in consideration of variable selection by applying to real plant data", The 27th International Symposium on Chemical Engineering (ISChE 2014), Kuala Lumpur, Malaysia, 6 December, 2014

## 6. 研究組織

### (1)研究代表者

柘植 義文 (TSUGE YOSHIKUMI)

九州大学・大学院工学研究院・教授

研究者番号: 00179988