

高炉操業における大規模データベース オンラインモデリング

伊藤 雅浩*・松崎 眞六*・大貝 晴俊*²・大舘 尚記*³・内田 健康*³
齋藤 信一*⁴・佐々木 望*⁵

Large Scale Database-based Online Modeling on Blast Furnace Operation

Masahiro ITO, Shinroku MATSUZAKI, Harutoshi OGAI, Naoki ODATE, Kenko UCHIDA, Shinichi SAITO and Nozomi SASAKI

Synopsis : With the advance of computer hardware performance and database system technology, it has been practicable to store and to search for a large quantity of data. So that Local modeling method, such as Just-In-Time modeling, has become very attractive in recent years. We have developed "Large scale database-based Online Modeling" as the practical method based on the Just-In-Time modeling concept on blast furnace operation, which has very complicated physical phenomena and strong non-linear specific characteristics. The validity of the developed modeling method has been confirmed by the study with blast furnace operation data, then the past similar operation data have been searched and the prospective operation data have been estimated very quickly and precisely.

Key words: ironmaking; modeling; computer; process control; production control; quality control; phase analysis.

1. 緒言

一般に、工学、自然科学、社会科学において、入出力を有するシステムを定義し、システムについて予測、制御、計画、設計、評価といった課題を数理的に考えるとき、目的に応じて適切なモデルを構築してシステムの特性を定量的に捉える必要がある。対象とするシステムの物理現象やメカニズムが予め正確に把握できる場合は、運動方程式やエネルギー保存則等の物理則に基づく物理モデリングが有効である。しかし、物理現象が数式で取扱うには複雑であるか全く未知のシステムに対しては、観測したデータ群が有する情報からシステムの特性を再現するモデリング手法が求められる。このようなシステムの入出力は強い非線形特性を有する場合が多いことから、非線形システムのモデリング問題として取り扱われている。

これまで、非線形システムのモデリングは、Artificial Neural Network(ANN)やHamilton-Jacobi-Bellman方程式等の大域的なモデリング手法、或いはGain-SchedulingやLocal Model Network(LMN)等の線形化を前提とした局所的なモデルを複数組み合わせる手法が研究されてきた。入出力を大域的に取扱う前者は、モデル構造が複雑化しかつ同定すべきモデルパラメータ数が多く実プロセスへの適用には課題が多い。また、線形化を前提とした後者は、定格設

計点以外の幅広い動作範囲における複数の平衡点に関する情報が予め必要であり、実プロセスでこのような情報を事前に把握しておくことは困難である。

一方、近年、計算機ハードウェアやデータベースシステム技術の発展に伴って大量データの蓄積と高速検索が可能となったこと等を背景に、“Just-In-Time（以後、JITと略す）モデリング”^{1,2)}或いは“Lazy Learning”^{3,4)}と呼ばれる新しい考え方の局所的なモデリング手法が注目を集めている。これらは、定格設計点以外の幅広い動作範囲において観測したデータをそのままデータベースとして蓄積しておき、システムの予測等の必要が生じる都度、入力である“要求点(Query)”に関連性の高いデータをデータベースから“近傍”データとして検索し、検索したデータの出力を補間する局所モデルを構成して“要求点”の出力を得るモデリング手法である。局所モデルを都度廃棄して観測データの更なる蓄積に対応していく点に特徴を有する考え方である。

一般に、ある時刻においてシステムから観測されるデータ、すなわちシステムの状態変数の一組（データセット）をシステムの相（または位相）と呼び、システムがとりうる相の全体をシステムの相空間(Topological Space)と呼ぶ。

このときシステムの相が n 個の数値の組で表わせるとき、 n をシステムの次元と呼び、 n 次元システムの相空間

平成16年2月24日受付 平成16年6月1日受理 (Received on Feb. 24, 2004; Accepted on June 1, 2004)

* 新日本製鐵(株) 環境・プロセス研究開発センター (Environment & Process Technology Center, Nippon Steel Corp., 20-1 Shintomi Futtsu 293-8511)

*² 早稲田大学大学院情報生産システム研究科 (Graduate School of Information, Production and Systems, Waseda University)

*³ 早稲田大学理工学部 (School of Science and Engineering, Waseda University)

*⁴ (株) 日鉄エレクトックスシステムソリューション事業部 (System Solution Div., Nittetsu Elex Corp.)

*⁵ 新日本製鐵(株) 名古屋製鐵所 (Nagoya Works, Nippon Steel Corp.)

は n 次元ユークリッド空間 R^n またはその一部の領域 D である⁵⁾。

ある時刻のシステムの相は、相空間 T 上の点である。このことを強調するために相のことを相点とも呼ぶ。

JITモデリングや Lazy Learning では“要求点”と観測データとの近傍を取扱うにあたり、システムの相空間上における状態変数の一組、すなわち相点の相互関係を定量的に評価する。定性推論の分野における事例ベース推論⁶⁾も同様の考え方に基づいている。

また、一般に、システムに非線形性が存在すると、例えば次数 n が小さいシステムであってもシステムの時間的変動(動的挙動, Dynamics) は複雑な挙動を呈する。

この場合、システムの相空間に観測データの時間遅れ座標軸を考慮すると、システムの時間的変動を顕在化できることがあり、このような相空間を非線形システム論では再構成状態空間と呼ぶ。本報では、以後、観測データの時間遅れ変数を“遅れ変数”と呼ぶ。

JITモデリングにおける課題は、“要求点”に類似した“近傍”データセットを検索するために、全ての観測データについて“要求点”との距離を求めて順序付ける処理を、システムの予測等を実施する都度に行う必要があるという点にある。高炉は物理現象が複雑かつ非線形性が強いプロセス(システム)であるため数多くのセンサが設置されており、観測データだけでなく、観測データの遅れ変数も考慮した相空間を用いてデータベースを構築しようとする場合、相空間の次数が増大してデータベースが大規模化し、その結果、“近傍”探索の計算負荷が非常に大きくなってオンライン適用が困難となる。

そこで、本報は、大規模なデータベースにJITモデリングをオンラインで適用するにあたりこのような問題を回避して実用となる手法を検討した。

本手法は、1)ステップワイズ法を応用して、観測データとこれらの遅れ変数から構成される変数群の中からシステムの出力に寄与する変数を選択してノイズとなる変数を積極的に削除し、2)選択された変数で構成される相空間の観測データセットを量子化した検索用データベースに格納しておき、3)“要求点”の“近傍”データセットを量子化された検索用データベース上で量子単位での検索を行うことで検索の効率化と計算負荷の大幅な低減を図り、4)検索した“近傍”データセットの出力を補間する局所モデルを用いて“要求点”の出力を推定する手法である。このとき、局所モデルは都度廃棄して観測データの更なる蓄積に備え、対象システムの経時的な特性変化に対応していく点は、JITモデリングと同じである。

本報では、この手法を“大規模データベースオンラインモデリング(LOM: Large scale database-based Online Modeling)”と呼び、本手法の有効性を検証するために高炉操業に適用した事例を報告する。

2. Just-In-Time モデリング^{1,2)}

大規模データベースオンラインモデリング(LOM)の基本的な概念となるJITモデリングを以下に説明する。

2.1 定式化

対象とするシステムは非線形で動的なシステムであり、次式のような回帰式モデルで与えられるとする。

$$y(t+p) = f\{y(t), y(t-1), \dots, y(t-n_y), u(t-d), u(t-d-1), \dots, u(t-d-n_u)\} \dots \dots \dots (1)$$

ここで、

- $u(t)$: 時間 t におけるシステムの制御入力ベクトル
- $y(t)$: 時間 t におけるシステムの観測出力ベクトル
- n_u : 制御入力ベクトルの次数
- n_y : 観測出力ベクトルの次数
- p : 予測時間
- d : むだ時間
- f : 未知の非線形関数
- t : 時間

である。式(1)で制御入力ベクトルは n_u 次までの遅れ変数を、観測出力変数ベクトルは n_y 次までの遅れ変数を考慮している。

ここで、システムの入力ベクトル x^k と出力ベクトル y^k を式(2)、式(3)のように再定義すると、

$$y^k = y(k+p) \dots \dots \dots (2)$$

$$x^k = \{y(k), y(k-1), \dots, y(k-n_y), u(k-d), u(k-d-1), \dots, u(k-d-n_u)\} \dots \dots \dots (3)$$

時間の推移に伴い入力ベクトル x^k と出力ベクトル y^k のデータセットが $(x^1, y^1), (x^2, y^2), \dots$ の如く対象とするシステムから大量に得られ、データ集合 $\{(x^k, y^k)\}, (k=1, 2, \dots)$ としてデータベースに蓄積される。 k は離散化時間である。このとき、JITモデリングは、予測や制御の要求の都度、蓄積されているデータ集合 $\{(x^k, y^k)\}$ から非線形関数 f を求めることに相当する。

すなわち、JITモデリングでは、例えば、システムの予測が必要となったとき、システムの現在の状態 $\{(x^k, y^k)\}$ を“要求点(Query)”とし、“要求点”に類似した“近傍”のデータセット $\{(x^k, y^k)\} (k < k_0)$ が過去に観測されてデータ集合に存在すれば、現在のシステムの時間的发展を記述する(すなわち予測する)非線形関数 f^{k_0} は、過去の非線形関数 f^k と似たものになるという考え方である。“近傍”データセットが複数存在すれば、これらのデータセットの出力 y^k を補間する局所モデルを用い、システムの出力 y^{k_0} を予測する。

予測後、都度、局所モデルは廃棄され、観測データが新たに蓄積されてデータ集合 $\{(x^k, y^k)\}$ が更新されていくこと

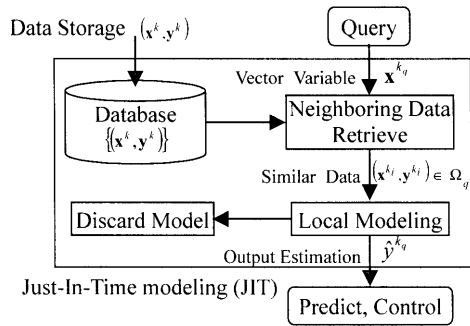


Fig. 1. Just-In-Time modeling.

で、対象プロセスの経時的な特性変化が次回の予測等に反映される。

2.2 “近傍”

JITモデリングでは“要求点”の“近傍”データセットを如何に正確に選択できるかが課題となる。データ相互(x^k, x^k)の近傍度を2つのベクトル間距離の大小で考える。距離d(k_i, k_j)としては、以下が提案されている¹⁾。

1) ユークリッド距離

$$d(k_i, k_j) = \sqrt{(\mathbf{x}^{k_i} - \mathbf{x}^{k_j})^T (\mathbf{x}^{k_i} - \mathbf{x}^{k_j})} \dots\dots\dots (4)$$

2) 重みつきユークリッド距離

$$d(k_i, k_j) = \sqrt{(\mathbf{x}^{k_i} - \mathbf{x}^{k_j})^T \mathbf{S} (\mathbf{x}^{k_i} - \mathbf{x}^{k_j})} \dots\dots\dots (5)$$

ただし、Sはスケーリング行列

3) Lpノルム (Minkowski 距離)

$$d(k_i, k_j) = \left(\sum_l |x_l^{k_i} - x_l^{k_j}|^p \right)^{\frac{1}{p}} \dots\dots\dots (6)$$

ただし、x_lは変数ベクトルxの第l成分

これらの距離関数を用いて、“要求点”との距離が近いデータセットを“近傍”として選択する。代表的な近傍選択方法には、k-NN(k Nearest Neighbors)法⁷⁾、k-BN(k Bipartite Neighbors)法⁸⁾、k-SBN(k Similar Bipartite Neighbors)法等、いくつかの手法が提案されている¹⁾。

2.3 局所モデル

JITモデリングにおける代表的な局所モデルには重み付き線形平均法(LWA)⁹⁾、重みつき局所回帰法(LWR)¹⁰⁾が提案されている¹⁾。

1) 重み付き線形平均法(LWA)

$$\hat{\mathbf{y}}^{k_q} = \frac{\sum_{i=1}^m w_i \mathbf{y}^{k_i}}{\sum_{i=1}^m w_i} \dots\dots\dots (7)$$

2) 重み付き局所回帰法(LWR)

$$\hat{\mathbf{y}}^{k_q} = (\mathbf{x}^{k_q})^T \hat{\boldsymbol{\theta}} \dots\dots\dots (8)$$

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^m w_i |\mathbf{y}^{k_i} - \hat{\mathbf{y}}^{k_i}|^2 \dots\dots\dots (9)$$

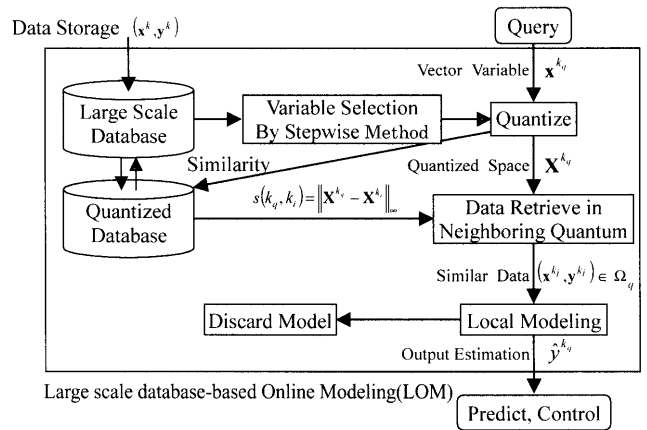


Fig. 2. Large scale database-based online modeling.

ただし、 $\hat{\mathbf{y}}^{k_q}$ は“要求点”ベクトル \mathbf{x}^{k_q} に対する出力の推定値ベクトルであり、 $\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}$ は局所モデルのパラメータとその同定値である。 w_i は“要求点”の“近傍”データセットのうち第i番目のデータに対応する“重み”であり、 m は“近傍”データセットの数である。

ここで、“重み”は“要求点”と“近傍”データセット間の距離dによって両者の距離が遠ければ0に近づき、逆に近ければ1に近づくように設定する。以下にJITモデリングで代表的な重み関数を示す¹⁾。

a) Gaussian関数

$$w(d) = e^{-d^2} \dots\dots\dots (10)$$

b) Tricube関数¹¹⁾

$$w(d) = \begin{cases} (1-d^3)^3 & \text{if } |d| < 1 \\ 0 & \text{otherwise} \end{cases} \dots\dots\dots (11)$$

c) 逆距離関数

$$w(d) = \frac{1}{1+d^p} \quad (p \text{ は正の整数}) \dots\dots\dots (12)$$

3. 大規模データベースオンラインモデリング (LOM)

大規模データベースオンラインモデリング(LOM)は、JITモデリングを実プロセスデータの相空間に適用するにあたり、ステップワイズ法による相空間の低次元化と、低次元化した相空間の量子化による近傍検索の効率化と、要求の都度生成し使用後廃棄される局所モデルで構成される。

3.1 ステップワイズ法による相空間の低次元化

ステップワイズ法は、回帰式モデルにおいて、できるだけ入力変数の数を少なくし、かつ観測値と予測値の差の平方和(残差平方和)が実用に耐え得るほど小さいものとするために、ある検定基準を設けて入力変数の追加、除去を行う方法である¹²⁾。

すなわち、ある入力変数を回帰式モデルに追加した場合、残差平方和の変化量を残差分散で正規化した値、いわゆる“変数の寄与率 F ”があらかじめ設けた検定基準より大きければその入力変数を追加し、ある入力変数を回帰式モデルから除去した場合の“変数の寄与率 F ”が検定基準より小さければ、その入力変数を除去する。この手順を出力変数との単相関係数の最も大きい入力変数から順に行ない、ある段階で追加される入力変数も除去される入力変数もなくなったとき、最終的に得られた回帰式を最良の回帰式とするものである。

高炉のように物理現象が複雑で非線形性の強い実プロセスにJITモデリングを適用するには、観測データのみならず観測データの遅れ変数を考慮した相空間の採用が重要である。高炉の場合、観測データの測定項目自体が膨大な数であり、これらの遅れ変数まで考慮すると極めて大きな次元の相空間を取扱わなければならない。

そこで本手法では、ステップワイズ法を応用し、厳しい検定基準を用いて相空間の低次元化を図ることで実用に供するモデリングを検討した。

3.2 相空間の量子化と“近傍”の検索

LOMでは、大規模かつ大量データをオンラインで高速に扱うことを実現するため、相空間上の観測データを量子化された検索用データベースに格納しておき、この量子化データベース上で“要求点”の“近傍”データセットを量子単位で検索することによって検索の効率化と計算負荷の大幅な低減を図る。

まず入力変数ベクトル \mathbf{x}^k の値域である入力空間を量子化し、量子空間 \mathbf{X}^k を定義することで入力変数ベクトル \mathbf{x}^k の分類を行う。

$$\mathbf{X}^k = Z(\mathbf{x}^k), \quad (i=1, 2, \dots, n) \dots\dots\dots(13)$$

ここで、 $Z(\cdot)$ は量子化演算子、 n は同一量子空間 \mathbf{X}^k に属するデータの数とする。このとき、量子化によって量子空間 \mathbf{X}^k は、入力変数ベクトル \mathbf{x}^k が1次元では区間、2次元では長方形、一般には超直方体となる。

ここで、量子空間 \mathbf{X}^k と \mathbf{X}^j との相似度 $S(k, j)$ を

$$S(k, j) = \|\mathbf{X}^k - \mathbf{X}^j\|_\infty \dots\dots\dots(14)$$

と定義する。ただし、 $\|\cdot\|_\infty$ は ∞ ノルムである。

このとき“要求点”ベクトル \mathbf{x}^{k_q} を含む量子空間を \mathbf{X}^{k_q} とし、“要求点”ベクトル \mathbf{x}^{k_q} の近傍空間 Ω_q を

$$\Omega_q = \left\{ \mathbf{X}^{k_p} \mid S(k_q, k_p) = \min_{\mathbf{X}^{k_p} \in T} S(k_q, k_p) \right\} \dots\dots\dots(15)$$

と定義する。ただし、 T は相空間を表わす。

量子化を導入することにより相似度 S は離散値となり、“近傍”を検索するには、まず“要求点”を含む同一量子、隣の量子、 \dots 、というように量子化データベース上で単純かつ効率的に検索することができる。

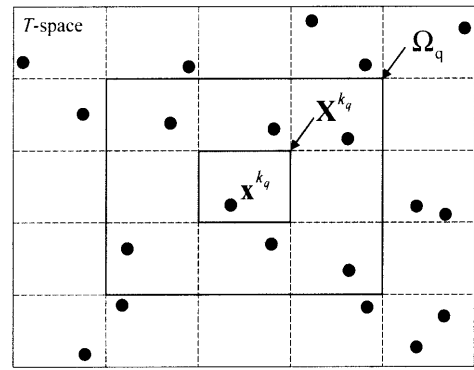


Fig. 3. Neighboring quantized space of query.

入力空間の量子化幅の決定方法については、いくつかの方法が考えられる。ここでは、高炉操業データへの適用の第1ステップとして、最も単純な一様均等分割法を用いた。一様均等分割法とは相空間の座標軸を各軸共通の量子化数で均等に分割する方法である。また量子化数は、モデル精度評価用データと推定値の関係を評価する Leave-one-out Cross Validation 法¹³⁾やスタージェスの公式¹⁴⁾によって決定した。

3.3 局所モデル

“要求点”ベクトル \mathbf{x}^{k_q} が与えられたとき、“要求点”の“近傍”データセットと局所モデルを用いて出力の推定を行う。

JITモデリングにおける代表的な局所モデルには、第2章で述べたように重み付き線形平均法(LWA)や重み付き局所回帰法(LWR)などが提案されているが、ここでは高炉操業データへの適用の第1ステップとして、最も単純な相加平均法を用いた。

すなわち、“要求点”ベクトル \mathbf{x}^{k_q} に対する出力の推定値ベクトル $\hat{\mathbf{y}}^{k_q}$ を、

$$\hat{\mathbf{y}}^{k_q} = F(\mathbf{x}^{k_q}) = \frac{1}{M} \sum_{\mathbf{y}^k: (\mathbf{x}^k, \mathbf{y}^k) \in \Omega_q} \mathbf{y}^k \dots\dots\dots(16)$$

で算出する。ここで、 M は近傍空間 Ω_q に属する出力ベクトル \mathbf{y}^k の個数である。

4. 高炉操業における大規模オンラインデータデータベースモデリング (LOM)

高炉には、温度、圧力、組成をはじめ各種物理量の検出端が多数(1千点以上)設置されている。各計測値はオンラインでプロセスコンピュータ(以後、プロコンと略す)に伝送され、蓄積される。また、これらの計測値を組み合わせ、例えば、炉内通気性指標である“K値”や炉内反応や炉体冷却状態の指標である“熱負荷”といった操業管理値がプロコン内で逐次算出され、プロコン内に蓄積される。

これらプロコン内に蓄積される大量の操業データをプロ

コンから取り出して大規模データベースを構成し、第3章で述べた大規模データベースオンラインモデリング(LOM)を適用して、過去の類似操業事例の検索と検索結果に基づく将来予測を実施し、本手法の有効性を検証した。

4.1 高炉操業データの大規模データベース

本報では、名古屋製鐵所第3高炉の操業データを対象データベースとした。データ項目には操業管理上重要とされる145項目を選択し、サンプリング時間は1時間とした。データ収集期間は、2003年4月1日~2004年1月28日でデータ点数は7266点である。

4.2 回帰式モデルの定式化

本報では、高炉プロセスを多変数入力・多変数出力系の非線形な動的システムとして捉える。このとき、高炉プロセスは複雑な物理現象が多数連成して成立しており、現象論的な解釈から“原因としての入力変数”と“結果としての出力変数”をあらかじめ明確に分離することは難しい。

そこで本報では、式(1)において入力変数ベクトル \mathbf{u} と出力変数ベクトル \mathbf{y} を等価に取り扱い、すべての変数は高炉から収集する変数ベクトル \mathbf{y} で記述できるものとする。すなわち、 p 時間後の高炉プロセスのダイナミクスは、式(17)のような回帰式モデルで表現できると仮定する。

$$\mathbf{y}(t+p) = f\{\mathbf{y}(t), \mathbf{y}(t-1), \mathbf{y}(t-2), \dots, \mathbf{y}(t-n_y)\} \dots\dots\dots(17)$$

ここで、式(17)を式(2)、式(3)の如く、

$$\mathbf{y}^k = \mathbf{y}(k+p) \dots\dots\dots(18)$$

$$\mathbf{x}^k = \{\mathbf{y}(k), \mathbf{y}(k-1), \mathbf{y}(k-2), \dots, \mathbf{y}(k-n_y)\} \dots\dots\dots(19)$$

と再定義したとき、高炉から得られた大量のデータセット $\{\mathbf{x}^k, \mathbf{y}^k\}, (k=1, 2, 3, \dots)$ がデータベースに蓄積される。

4.3 ステップワイズ法を応用した相空間の低次元化

高炉から収集した変数の数が N 個で、このうち第1変数 y_1 を回帰式モデルの出力として式(17)を変数ベクトルの各要素で書くと式(20)となる。

$$y_1(t+p) = f \left\{ \begin{matrix} y_1(t), & y_1(t-1), & \dots, & y_1(t-n_1), \\ y_2(t), & y_2(t-1), & \dots, & y_2(t-n_2), \\ \dots, & \dots, & \dots, & \dots, \\ y_N(t), & y_N(t-1), & \dots, & y_N(t-n_N) \end{matrix} \right\} \dots\dots\dots(20)$$

式(20)は p 時間後の第1変数の値 $y_1(t+p)$ が $\sum_{i=1}^N (n_i+1)$ 個の変数の回帰式で表現されることを示している。

高炉操業データでは観測する変数の数 N が大きく、また短い周期から長い周期の特性を有する変数が多数混在しているため次数 n_y の値もあらかじめ大きく設定する必要があるため、その結果、回帰式を構成する変数の数が膨大な数となる。

例えば、 $p=1$ とし予測したい出力変数 y_1 を溶銑温度とする。1時間後の溶銑温度に影響を与える変数因子として経験的に現在値を含む過去12時間以内のデータで表現でき

Table 1. Selected variables for molten iron temperature after 1 h.

Selected variables by stepwise method		F value
Number	Content	
1	Molten iron temperature (present)	21225
2	Molten iron Si concentration (before 1 hour)	496
3	Molten iron Ti concentration (present)	273
4	Tapping velocity (present)	215
5	Molten iron temperature (before 4 hours)	104
6	Molten iron temperature (before 1 hour)	87.3
7	Top gas CO concentration (present)	76.0
8	Top gas CO concentration (before 1 hour)	71.3
9	Pulverized Coal Ratio (present)	65.2
10	Taphole number (before 2 hours)	65.1
...
32	Slag Al2O3 concentration (before 2 hours)	20.4

ると仮定、すなわち $n_1=n_2=\dots=n_{145}=12$ とすると、1時間後の溶銑温度 $y_1(t+1)$ は、式(21)の如く1885個の変数を持つ回帰式モデルとなる。このとき、式(21)で記述されるシステムの相空間の次数は1885次元である。

$$y_1(t+1) = f \left\{ \begin{matrix} y_1(t), & y_1(t-1), & \dots, & y_1(t-12), \\ y_2(t), & y_2(t-1), & \dots, & y_2(t-12), \\ \dots, & \dots, & \dots, & \dots, \\ y_{145}(t), & y_{145}(t-1), & \dots, & y_{145}(t-12) \end{matrix} \right\} \dots\dots\dots(21)$$

ここで、LOMでは、ステップワイズ法を用いて1時間後の溶銑温度 $y_1(t+1)$ に対する変数の寄与率 (F 値) を計算して入力変数の追加及び除去を実施した。追加検定基準 F_{in} と除去検定基準 F_{out} に一般に検定基準として採用される $F_{in}=F_{out}=2.0$ と設定した場合、ステップワイズ法は261変数を選択した。しかし261変数のままではノイズとなる変数がまだ多く存在し、その結果、溶銑温度の予測精度が十分でなかったため、最終的に F 値が20以上となる上位32変数を選択した(Table 1)。

すなわち、1時間後の溶銑温度 $y_1(t+1)$ を精度よく予測するシステムとして、あらかじめ仮定した1885次元の相空間を32次元の相空間へと低次元化した。

4.4 相空間の量子化とLOMによる予測精度の検証

ステップワイズ法を応用して選択した遅れ変数を含む32変数について、各々量子数20で量子化し、量子化した32次元の相空間を構築した。量子化数の設定にあたってはいくつかの指針があるが、ここではスタージェスの公式によって得られる量子数、すなわち、

$$N_q \cong 1 + \log_2 n_{all} = 1 + \log_2 7266 \cong 14 \dots\dots\dots(22)$$

や Leave-one-out Cross Validation 法を用いて何通りか設定し、溶銑温度の予測精度が最良となる量子数20を選択した。溶銑温度の予測精度は、全データ7266点のデータセットからランダムに200セットを取り出してそれぞれを“要求点” \mathbf{x}^k とし、LOMによる1時間後の溶銑温度の予測

値 y_1^{k+1} と実績値 y_1^k との相関で評価した。このとき相関係数 ρ は 0.739 であり、1 時間後の溶銑温度が良好に予測できることが確認できた (Fig. 4)。

4.5 LOM による過去事例検索と将来予測事例

大規模データベースの中から任意のデータセットを“要求点”ベクトル x^k とし、量子化された相空間上で“近傍”にある量子内に存在するデータセットを過去のデータセットの中から選択することで、“要求点”と類似する過去の操業事例を効率よく検索することができる。

例えば、2003年4月1日～2004年1月28日の全データ

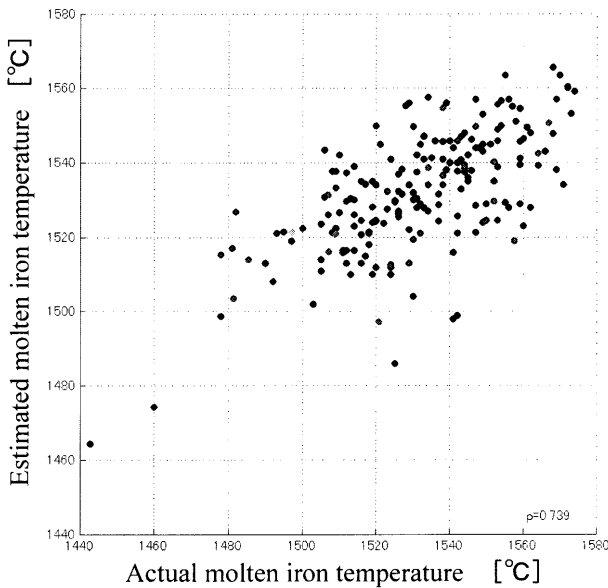


Fig. 4. Correlation between 1 h later actual molten iron temperature and estimated molten iron temperature by LOM.

セット 7266 点の中から 2004 年 1 月 1 日 6:00 のデータセットを取り出して“要求点”とし、“要求点”より過去のデータセットから“近傍”データセットを検索すると、相似度 $S=2$ の近傍量子内に 2 つのデータセットが存在し、2004 年 1 月 1 日 6:00 と似た操業状態が過去に 2 ケース存在したことがわかる (Fig. 5(b))。

このとき、相似度 $S=0,1$ すなわち同一量子化空間や一つ隣の量子には類似な操業事例が存在しなかった。

LOM は、次元 32、量子数 20 の量子化空間の中から、まず相似度 $S=0$ すなわち“要求点”が属する量子内で過去データの有無をチェックした。同一量子内には過去データが無かったため、次に相似度 $S=1$ の量子内で過去データの有無をチェックした。相似度 $S=1$ の量子にも過去データが無かったため、次に相似度 $S=2$ の量子内で過去データの有無をチェックしたところ、過去データが 2 ケース存在した。

以上のように、“要求点”との距離を全データ 7266 点に対して 7266 回計算して求めて順序付けしなければならないところを、LOM では対象量子における過去データの有無を 0 (無) 又は 1 (有) のフラグをチェックするだけで完了し、かつ相似度 S の値を離散的に順を追って量子を選択してその有無をチェックするだけで“要求点”と類似する過去の操業事例を検索することができるので個々の過去データとの距離計算をする必要が無く、計算量を大幅に削減することができる。

今回の事例では、LOM は Table 1 で示した遅れ変数を含む 32 次元の相空間上の 2003 年 4 月 1 日～2004 年 1 月 28 日の全相点 7266 点の中から、2004 年 1 月 1 日 6:00 に至るまでの

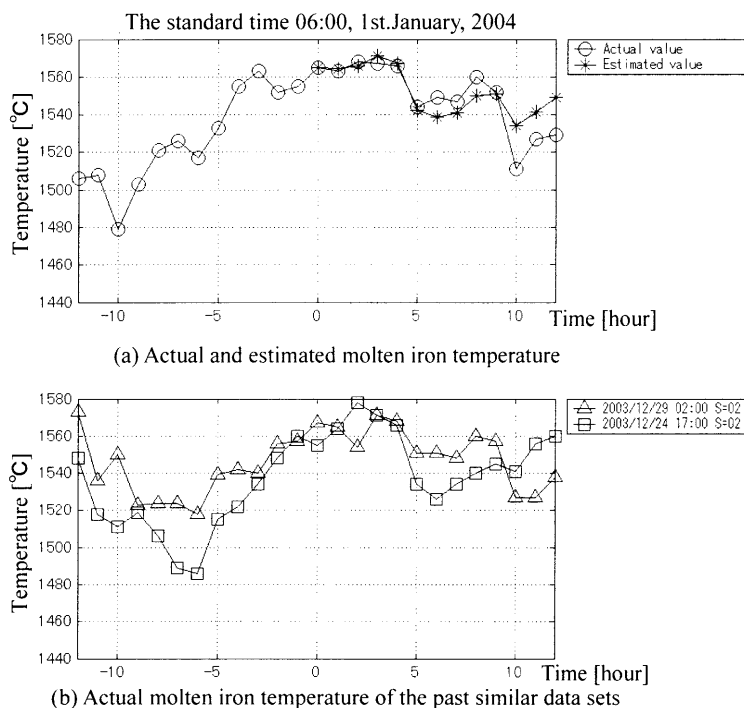


Fig. 5. The past similar data sets and estimated result of molten iron temperature by LOM.

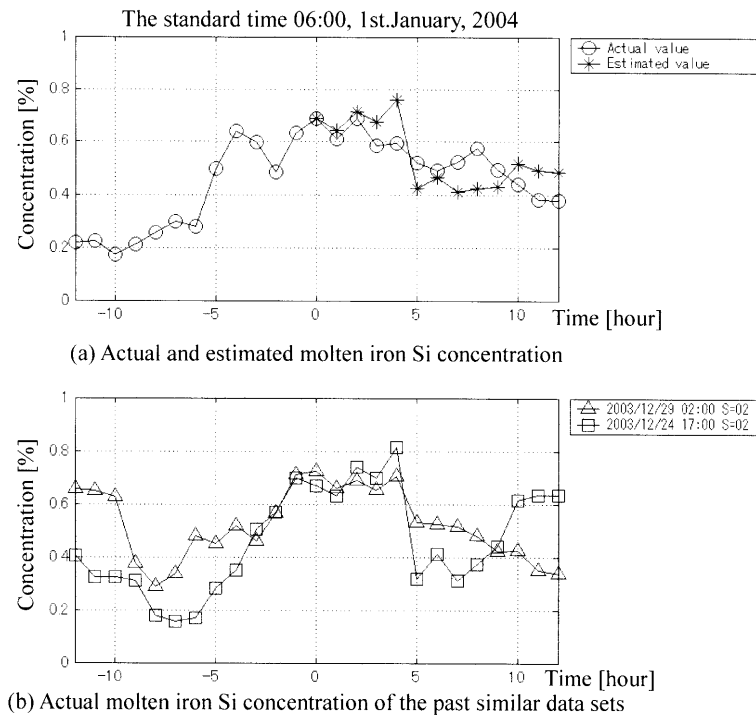


Fig. 6. The past similar data sets and estimated result of molten iron Si concentration by LOM.

高炉の時間的変動を記述する“要求点”の“近傍”に存在する、2003年12月24日17:00および2003年12月29日02:00の2つの相点を検索した。

すなわち、これら2つの相点が記述する高炉の時間的変動は、高炉の非線形特性を考慮した上で相似度 $S=2$ の範囲で2004年1月1日6:00に至るまでの時間的変動と類似であると判断できる。

検索された2つの“近傍”データセットの出力ベクトル y^k に対し、式(16)の相加平均法を用いて出力変数、すなわち1時間後の溶銑温度を推定し、その値をFig. 5(a)に示す。Fig. 5(a)からステップワイズ法で変数の寄与率を計算するさいに出力変数と定義した1時間後の溶銑温度が精度良く予測できていることが確認できた。

さらに、この2つの“近傍”データセットの1時間後以降の実績値を用いて、1時間後だけでなく基準時間から将来の12時間にわたる溶銑温度の推移を式(16)を用いて予測してみると、2004年1月1日6:00以降12時間の溶銑温度推移の実績値と予測値は良く一致することが確認できた(Fig. 5(a))。

すなわち、1時間後の溶銑温度の予測に対する変数の寄与率に着目して相空間を構成する変数を選択したが、過去の類似事例を相似度 $S=2$ の量子内で検索できれば、1時間後以降の溶銑温度の推移も十分予測可能であることが確認できた。

さらに、高炉を多変数出力系のシステムとして捉えた場合、式(16)で算出する出力の推定値ベクトル \hat{y}^k の他の変数要素、例えば、溶銑Si濃度(Fig. 6)や K 値(Fig. 7)を取り出してみると、溶銑温度と同様にこれらの過去の“近傍”

データセット検索と1時間後および1時間後以降の将来予測も可能であることが確認できた。

以上のように、LOMを用いることによって短時間で精度良く過去の類似事例を検索しその日時を特定することができ、操業日誌からそのときの操業アクションを確認すると共にその後の操業状態の推移を予測することで、今後の操業アクションの判断にあたって操業者に有益な情報を提供する操業支援技術として、その有効性が確認できた。

5. 結言

本報は、計算機ハードウェアやデータベースシステム技術の発展に伴って大量のデータの蓄積と高速検索が可能となったこと等を背景に、近年、注目されているJITモデリングを、物理現象が複雑で非線形性の強い高炉にオンライン適用するにあたり実用的な手法を検討し、高炉実操業データを用いてその有効性を確認した。このとき、過去の類似事例検索、将来予測に関する計算は極めて短時間に実行され、オンライン適用が十分可能であることを確認した。

また、高炉操業データは現在逐次蓄積中であり、定格設計点以外の幅広い動作範囲で観測データがそのままデータベースとして蓄積されることにより、幅広い操業範囲での高炉の操業予測も可能であると考えられる。

本報では、手法の検討にあたり、その第1ステップとして、入力空間の量子化幅の決定には最も単純な一様均等分割法を用い、局所モデルにも最も単純な相加平均法を用いた。局所モデルにJITモデリングで既に提案されている距

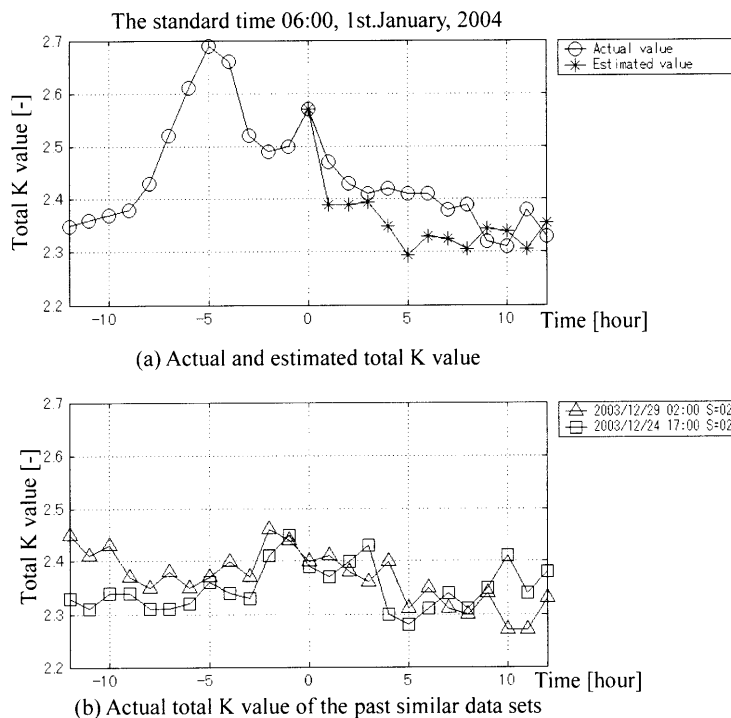


Fig. 7. The past similar data sets and estimated result of total K value by LOM.

離に基づく重み関数を用いたモデルを採用すること等により、更なる予測精度の向上が可能であると考えられる。

今後、局所モデルの最適化やデータセットの蓄積に伴うステップワイズ法による相空間の再低次元化（選択変数項目の更新や学習）の必要性可否を検討すると共に、1時間後の溶銑温度のみを評価関数とするのではなく多変数入力・多変数出力を有する非線形な動的システムに対する汎用的な手法へと改良していく予定である。

文 献

- 1) Q.Zheng and H.Kimura: *Trans. Soc. Instrum. Control Eng.*, **37** (2001), 640.
- 2) Q.Zheng and H.Kimura: *Asian J. Control*, **3** (2001), 35.
- 3) C.G.Atkeson, A.W.Moore and S.Schaal: *Artificial Intelligence Rev.*, **11** (1997), 11.
- 4) G.Bontempi, M.Birattari and H.Bersini: *Int. J. Control*, **72** (1999), 643.
- 5) 丹羽敏雄：微分方程式と力学系の理論入門，遊星社，東京，(1988)，31.
- 6) H.Tsutsui, A.Kurosaki and T.Sato: *Trans. Soc. Instrum. Control Eng.*, **33** (1997), 947.
- 7) J.Zhang, Y.Yim and J.Yang: *Artificial Intelligence Rev.*, **11** (1997), 175.
- 8) Q.Zheng and H.Kimura: 42nd Japan Joint Automatic Control Conf., The Society of Instrument and Control Engineers, Tokyo, (1999), 143.
- 9) D.Kibler, D.W.Aha and M.K.Albert: *Comput. Intelligence*, **5** (1989), 51.
- 10) C.G.Atkeson, A.W.Moore and S.Schaal: *Artificial Intelligence Rev.*, **11** (1997), 11.
- 11) W.S.Cleveland and S.J.Delvin: *J. Am. Stat. Assoc.*, **403** (1988), 83.
- 12) 河口至商：多変数解析入門Ⅰ，森北出版，東京，(1973)，27.
- 13) D.J.C.MacKay: *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, (2003), 353.
- 14) 統計学入門，東京大学教養学部統計学教室編，東京大学出版会，(1991)，22.