半教師あり学習を用いた PWARX システムの同定

牧 原 史 弥*・笠 嶋 丈 夫*・小 川 雅 俊*

Identification of PWARX Systems Using Semi-supervised Learning

Fumiya MAKIHARA*, Takeo KASAJIMA* and Masatoshi OGAWA*

This paper proposes method of identification of piecewise affine autoregressive exogenous systems using semi-supervised learning. The method enables a model designer to insert some class-labeled data to modeling scheme. The effectiveness of proposed method is ensured by numerical simulation.

Key Words: system identification, PWARX system, semi-supervised learning, sum-of-norms regularization

1. はじめに

本稿では、システムの状態や入力により動特性が切り替わるシステムのうち多項式で記述される PWARX (piecewise affine autoregressive exogenous) システムの同定問題を扱う。PWARX システムは複雑な非線形システムの近似として扱える点で実用的であり、その同定手法の研究は重要である。

Garulli ら¹⁾の調査によると、PWARX システムの同定手法はクラスタリングベースまたは最適化ベースのアプローチに分類される。前者は、教師なしクラスタリングにより同定用データをサブシステムごとに分類することを特徴とする方策である。たとえば、Nakada ら²⁾のガウス混合モデルを用いた同定手法や Tabatabaei-Pour ら³⁾の k-plane クラスタリングを再帰的に用いる同定手法がある。一方、後者のアプローチは PWARX システムの同定問題の一部を数理最適化問題へ定式化して解く方策であり、代表的な手法として Ohlsson ら⁴⁾の sum-of-norms 正則化を用いた手法がある。

これら先行研究で提案されている同定手法は広い用途で有用であるが、使用する同定用データによってはモデルの精度が十分に得られない場合があった。一例としては、データの不足により本来異なるサブシステムから生成されたデータ間の類似度が相対的に高まり、同一のサブシステムのデータとして誤認される場合である。これは、同定用データの類似度にのみ着目してクラスタリングや最適化を行なうことに起因する。

この問題に対して、本稿ではクラスタリングベースと最適化ベースの2種類の従来手法をそれぞれ拡張して、半教師あり学習を用いた PWARX システムの同定手法を提案する。半教師あり学習とは、正解ラベルのあるデータとラベルのないほかのデータを併用して分類モデルや回帰モデルを学習する機械学習の枠組みである。提案手法では半教師あり学

習を用いた同定を実施するにあたり、入出力データのほかに、一部の入出力データがどのサブシステムから生成されたかを示すラベルを使用する。たとえば、プラントによっては、毎分測定される入出力データのほかに、1日に数回しか観測されないプラントの状態を示す定性的なデータがあり、これらに便宜上の値を割り振ることによりラベルとして扱うことができる。ラベルの設定後、クラスタリングベースの提案手法では、ラベルありデータとラベルなしデータをもとに同定用データのクラスタリングを行ない、各クラスごとにサブシステムのパラメータを推定する。また、最適化ベースのアプローチでは、サブシステムのパラメータを推定する従来の最適化のスキームにラベルを組み込む手法を提案する。これらの拡張により、データ間の類似度のみでなくラベルに基づいた PWARX システムの同定が可能となり、特定の同定用データで生じるモデル精度の低下が抑制されることを期待できる。

2. PWARX システム

PWARX システムは、多項式を用いて動特性の切り替わりが表現されるシステムであり、(1) 式により記述される。

$$y(k) = \theta_i^T \varphi(k) + e(k) \quad \text{if } \varphi(k) \in \Pi_i \quad (1)$$

ここで、 $y(k)$ はシステムの出力、 $\varphi(k)$ はシステムの入出力を並べた回帰ベクトル、 θ_i はサブシステムのパラメータベクトル、 $e(k)$ は白色雑音とする。また、領域 Π_i は回帰ベクトルの定義域のうちサブシステム i に対応する部分集合とする。

PWARX システムの同定問題は、 N セットの入出力データ $\{y(k), u(k)\}_{k=1}^N$ が与えられたとき、モデル構造、モデル次数、サブシステム数、各サブシステムのパラメータベクトル、分離超平面を求める問題である。分離超平面とは、回帰ベクトルの定義域をサブシステム数の領域に分ける超平面である。一般的に、上記問題を一括に解くことは難しいため、本稿ではモデル構造、モデル次数、サブシステム数は既知とする。

3. 半教師あり学習による PWARX システムの同定

本章では、クラスタリングと最適化の2つのアプローチにおいて、半教師あり学習を用いた同定手法を提案する。

* (株) 富士通研究所 厚木市森の里若宮 10-1

* FUJITSU LABORATORIES LTD., 10-1 Morinosato-Wakamiya, Atsugi

(Received May 27, 2018)

(Revised September 30, 2018)

Fig. 1 Strategies of clustering-based identification

3.1 クラスタリングベースドアプローチ

クラスタリングに基づく従来の同定手法は、Fig. 1 上側に示される段階的な手順をとる。はじめに、同定用データセット $\{\varphi(k)\}_{k=1}^N$ に対して教師なしクラスタリングを行なう。つぎに、推定したクラスごとにサブシステムのパラメータ θ_i を最小二乗法により求める。最後に、任意の回帰ベクトルが属するクラスを判別する分類器を作成する。

これに対して、提案手法の手順を以下に記す (Fig. 1 下側)。簡単のため 2 クラスの場合を扱うが、一対他法をはじめとする多クラス分類の技法により 3 クラス以上へ容易に拡張できる。

はじめに、同定用データセット $\{\varphi(k)\}_{k=1}^N$ に含まれる一部のデータに対して、データを生成したサブシステムを示すラベルの紐付けを行なう。つまり、ある時刻 k_j におけるデータ $\varphi(k_j)$ に対してラベル $z(k_j) \in \{-1, 1\}$ を紐付ける操作をラベル数 M だけ行なう。これより、ラベルありデータセット $\{\varphi(k_j), z(k_j)\}_{j=1}^M$ およびラベルなしデータセット $\{\varphi(k_j)\}_{j=M+1}^N$ を生成する。

つぎに、ラベルありデータおよびラベルなしデータを用いた半教師あり学習により回帰ベクトルの分類器を作成し、分類器によりラベルなしデータのクラスを推定する。本稿では、分類器のモデルとして以下のカーネルモデル (2) を考える。

$$f_{\omega}(\varphi(k_j)) = \sum_{j'=1}^N \omega_{j'} K(\varphi(k_j), \varphi(k_{j'})) \quad (2)$$

$\omega_{j'}$ は分類モデルのパラメータ、 ω は $\omega_{j'}$ を並べたベクトル、 $K(\cdot, \cdot)$ は (3) 式で記述されるガウスカーネルとする。

$$K(\mathbf{x}, \mathbf{c}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2h^2}\right) \quad (3)$$

ただし、 $\mathbf{c}$ は重心、 h はバンド幅、 $\|\cdot\|$ は l_2 ノルムとする。いま、(2) 式のパラメータ ω は、データセットを用いて以下の半教師あり学習の評価式 (4) より算出される⁵⁾。

$$\begin{aligned} \min_{\omega} \frac{1}{2} \sum_{j=1}^M (f_{\omega}(\varphi(k_j)) - z(k_j))^2 + \frac{\lambda}{2} \|\omega\|^2 \\ + \frac{\nu}{4} \sum_{j=1}^N \sum_{j'=1}^N W(\varphi(k_j), \varphi(k_{j'})) \\ \times (f_{\omega}(\varphi(k_j)) - f_{\omega}(\varphi(k_{j'})))^2 \end{aligned} \quad (4)$$

ここで、 λ, ν は学習の重みパラメータとする。第 1 項はラベルありデータにより学習する項であり、第 2 項は過学習を抑制する l_2 正則化項である。第 3 項はラプラス正則化項と

いい、回帰ベクトル間の類似度 $W(\varphi(k_j), \varphi(k_{j'}))$ をもとに、ラベルなしデータを含むすべてのデータにより学習する項である。なお、この評価式の解は解析的に求まることが知られている。分類モデルの学習後は、符号関数 (5) によりラベルなしデータセットの各クラスを推定する。

$$\hat{z}(k_j) = \text{sign}(f_{\omega}(\varphi(k_j))) \in \{-1, 1\} \quad (5)$$

最後に、クラスごとにデータをまとめ、サブシステムのパラメータベクトル θ_i を最小二乗法により算出する。

3.2 最適化ベースドアプローチ

本稿では、最適化に基づくアプローチとして、現実的な計算コストで解ける sum-of-norms 正則化による同定手法を扱う⁴⁾。この手法では、同定用データを直接的に教師なしクラスタリングせず、以下の凸最適化問題 (6) を用いて各時刻で異なるパラメータベクトル $\theta(k)$ をサブシステム数と同数になるまで集約することにより、サブシステムのパラメータベクトル θ_i を推定する。その後、集約された θ_i をもとに回帰ベクトル $\varphi(k)$ の属するクラスを推定する分類器を生成する。

$$\begin{aligned} \min_{\theta(k), k=1, \dots, N} \sum_{k=1}^N (y(k) - \theta^T(k) \varphi(k))^2 \\ + \lambda \sum_{k=1}^N \sum_{k'=1}^N W(\varphi(k), \varphi(k')) \|\theta(k) - \theta(k')\| \end{aligned} \quad (6)$$

ここで、(6) 式の意味について簡潔に述べる。第 1 項は各時刻のパラメータ $\theta(k)$ を推定する項であり、この項だけでは過学習は必至である。第 2 項は、sum-of-norms 正則化項とよばれ、データの類似度により時刻ごとに異なる $\theta(k)$ を同値とするはたらきをもつ。この手法はパラメータ推定の最適化をクラスの推定に使用する点で画期的であるが、データ間の類似度のみに基づき $\theta(k)$ を集約しているため、使用する同定用データによってはモデルの精度が低下することがあった。

これより、以下の (7) 式による拡張手法を提案する。

$$\begin{aligned} \min_{\theta(k), k=1, \dots, N} \sum_{k=1}^N (y(k) - \theta^T(k) \varphi(k))^2 \\ + \lambda \sum_{k=1}^N \sum_{k'=1}^N W(\varphi(k), \varphi(k')) \|\theta(k) - \theta(k')\| \\ + \nu \sum_{j=1}^M \sum_{j'=1}^M \delta(z(k_j), z(k_{j'})) \|\theta(k_j) - \theta(k_{j'})\| \end{aligned} \quad (7)$$

ただし、 $\delta(\cdot, \cdot)$ はクロネッカーのデルタとする。追加された第 3 項はデータの類似度ではなくラベル $z(k)$ により $\theta(k)$ を集約するはたらきをもつ。これより、類似データが不足している場合や複数のサブシステムでデータが類似する場合に、モデル精度の低下を抑制する効果が期待できる。

4. 数値例

数値例により提案手法の有用性を確認する。対象は以下の (8) 式で表わされる飽和系の PWARX システムとする。

Fig. 2 Class-labeled data (circles, triangles, diamonds)

$$y(t) = \begin{cases} [a_1 & a_2 & 0 & -2]\varphi(t) + e(t) & [0 & 0 & 1 & 2]\varphi(t) < 0 \\ [a_1 & a_2 & 1 & 0]\varphi(t) + e(t) & [0 & 0 & 1 & 2]\varphi(t) \geq 0, \\ & [0 & 0 & 1 & -2]\varphi(t) \leq 0 \\ [a_1 & a_2 & 0 & 2]\varphi(t) + e(t) & [0 & 0 & 1 & -2]\varphi(t) > 0 \end{cases} \quad (8)$$

ただし, $a_1 = -0.5$, $a_2 = -0.1$ とする. 回帰ベクトルは, $\varphi(t) = [y(t-1) \ y(t-2) \ u(t-1) \ 1]^T$ であり, 入力 $u(t)$ は平均 0, 分散 4^2 のガウス分布より生成し, 雑音 $e(t)$ は平均 0, 分散 0.4^2 の白色雑音とする. 同定用および検証用データはともに 100 サンプルとする. 比較には従来の sum-of-norms 正則化による同定手法を使用し, 予測精度の指標には (9) 式の適合率⁶⁾を使用する.

$$\text{FIT} = \left(1 - \frac{\sqrt{\sum_{k=1}^N (\hat{y}(k) - y(k))^2}}{\sqrt{\sum_{k=1}^N (y(k) - \bar{y})^2}} \right) \times 100 \quad (9)$$

ただし, $y(k)$ は出力の真値, $\hat{y}(k)$ は出力の予測値, $\bar{y}$ は $y(k)$ の時間平均とする. 適合率は $-\infty$ から 100% までの値をとる.

ここで, 提案手法において使用したラベルを Fig. 2 に示す. 本検証では, サブシステムごとに 3 点のデータを無作為に選択し, おおのこのラベルを付与した. ただし, 実用の場合にはラベルが事前に存在することが前提であることに注意する.

同定の設定はバリデーションにより選択し, 以下のとおりとする. 従来手法では, 重みパラメータ λ は 0.01, 類似度関数は $k = 6$ の k 最近傍類似度 (近傍のとき 1, それ以外は 0 の値をとる) とする. 最適化ベースの提案手法では, 従来手法と同様の設定に加え, 重みパラメータ ν を 10 とする. これら 2 つの手法では, 分類器に決定木を用いる. また, クラスタリングベースの提案手法では, λ および ν はともに 1, 類似度関数にはバンド幅 1 のガウスカーネルを用いる. 数値計画ソフトには CVX⁷⁾を使用する.

以上の条件のもと, 実行したシミュレーションの結果を Fig. 3 に示す. ここで, Fig. 3 の 6 つの図のうち, 上段は従来手法, 中段は最適化ベースの提案手法, 下段はクラスタリングベースの提案手法の結果である. また, 左側の図は同定用データのクラスタリング結果およびシステム本来の分離超平面を表わし, 右側の図は検証用データに対するシミュレーション予測の一部を表わす. 予測精度である適合率は, 従来手法は 63.7%, 最適化ベースの提案手法は 85.5%, クラスタリングベースの提案手法が 87.7% であり, 提案手法による精度向上が確認された. いま, 従来手法の結果 Fig. 3 (b)

Fig. 3 Left figures show classified data and original separating hyperplanes. On the right, solid lines show model output and dot lines show true output (upper: previous method, middle: proposed optimization-based method, lower: proposed clustering-based method)

に着目すると, $y(k)$ が 0 以下のときに予測がずれていることがわかる. これは, Fig. 3 (a) において丸印と三角印のデータが適当に分類されなかったことが原因である. 一方, 提案手法である Fig. 3 (c) および (e) では従来手法に対して比較的良好に分類されており, 予測の誤差も小さいことがわかる.

5. おわりに

本稿では, 半教師あり学習を用いた PWARX システムの同定手法として, クラスタリングおよび最適化に基づく 2 つの手法を提案し, 数値例により有効性を確認した. 今後は, ラベル数や使用するラベルありデータ間の類似度が結果に与える影響を定量的に扱うことが研究が課題となる.

参考文献

- 1) A. Garulli, S. Paoletti and A. Vicino: A survey on switched and piecewise affine system identification, *Proc. 16th IFAC Symp. Syst. Identif.*, 344/355 (2012)
- 2) H. Nakada, K. Takaba and T. Katayama: Identification of piecewise affine systems based on statistical clustering technique, *Automatica*, **41**-5, 905/913 (2005)
- 3) M. Tabatabaei-Pour, K. Salahshoor and B. Moshiri: A k-plane clustering algorithm for identification of hybrid systems, *Proc. 6th World Congr. Intell. Contr. Automat.*, 1333/1337 (2006)
- 4) H. Ohlsson and L. Ljung: Identification of switched linear regression models using sum-of-norms regularization, *Automatica*, **49**-4, 1045/1050 (2013)
- 5) 杉山 将: イラストで学ぶ機械学習, 講談社 (2013)
- 6) 足立修一: システム同定の基礎, 東京電機大学出版局 (2009)
- 7) M. Grant and S. Boyd: The CVX Users' Guide Release 2.1, CVX Research (2015)