

SVM を用いた最適な学習変数予測手法

Method to predict process signals to learn using SVM

岡山大学
岡山大学

箕輪 弘嗣
五福 明夫

Hirotsugu MINOWA
Akio GOFUKU

岡山大学
岡山大学

Abstract

Study of diagnostic system using machine learning to reduce the incidents of the plant is in advance because an accident causes large damage about human, economic and social loss. There is a problem that 2 performances between a classification performance and generalization performance on the machine diagnostic machine is exclusive. However, multi agent diagnostic system makes it possible to use a diagnostic machine specialized either performance by multi diagnostic machines can be used. We propose method to select optimized variables to improve classification performance. The method can also be used for other supervised learning machine but Support Vector Machine. This paper reports that our method and result of evaluation experiment applied our method to output 40% of Monju.

Keywords: SVM (Support Vector Machine), learning machine, optimization, Monju, plant

1. 緒言

事故が起きると人的、経済的、社会的損失が大きい。そのため、プラントの事故を減らせるよう機械学習を用いた診断システムの高度化が進められている [1][2][3][4]。

その中でも注目を浴びているのが、Support Vector Machine(以下 SVM)である。SVM は、予めプロセス信号データとその時の運転状態との関係を学習した診断器を作ることができる技術である。様々な異常を学習した診断器を予め作っておき、運転時に用いることでいち早くプロセスの診断が可能となる。

しかし、SVM の性能改善には問題がある。SVM には分類性能と汎化性能が排他的であるため、一方の向上によって、もう一方の性能が低下してしまうおそれがある [4][5]。

問題を解決するため、マルチエージェント診断システムのフレームワークを適用する。diversity methodology [6] を基に作成された五福ら[7]のフレームワークでは、未知の異常を診断するエージェントと、個々の異常の発見に特化したエージェント

とを分離する事で、汎化性能、分類性能各々を独立して向上させることができるようになった。また、分離した事によって各診断エージェントで異なる機械診断手器を適用することができる利点がある。機械学習は診断には不要信号の影響を抑える機能が有るが、ノイズの影響を受けやすい手法もあるため、モデルスコアという機構を設け、予め有益な信号化否かを判断し、訓練データと評価データを用いた指数により有限時間に解析が終わる準最適解を求める手法の研究を行った。

本論文では、提案手法と、その手法をもんじゅ 40%出力時のシミュレータデータへ適用した結果を報告する。

2. 最適化法

2.1 提案内容

SVM から機械学習はノイズを除去できる機構が備わっているが、診断に有益な情報か否かを判断することはできないため、ノイズの影響を完全に除去できるわけではない。そこでモデルスコアなる信号の有益性を評価する指標を用い機械学習の弁別結果に応じて予め不要な信号を除去することで影響を抑える事を試みた。

機械学習では複雑な非線形関数を作成するため、変数を組み合わせた時の精度の変化は厳密に推測することは

連絡先：箕輪弘嗣, 〒700-8530 岡山市北区津島中
3-1-1, 岡山大学大学院自然科学研究科,
E-mail : minowa@sys.okayama-u.ac.jp

困難であるが、大まかな予測は可能でないかと考えた。精度が向上するためには、データの公平分布性や正確性が必要であり、運転状態を判断する上で、信号群がそれぞれを補う形で空間を形成できる必要がある。公平分布性であれば説明変数が取りうる空間を網羅できるか否かであるが、閉じた空間で変数の組合せは、少数の変数組合せを掛けあわせて表現できる。網羅性があるため、少ない変数の組み合わせでも、学習した結果を元に正答率の増減の幅を算出できるのではないかと考えた。例えば、変数 A, B, C の組合せにおける正答率を算出したいが、(A, B, C) を学習できなくても、(A, B) と (B, C) と (C, A) の組合せにおける関係性から (A, B, C) の組み合わせを学習した時の正答率を近似できるというものである。

本手法は閉じられた空間における変数と正答率の関係を機械診断器で予測する事で最適な学習変数を見つけることにした。

2.2 モデルスコア

本業務では、異常徴候検出を行う SVM の検出性能を評価するため、状態判別の正常時と異常時の正答率の平均値を用いて、異常徴候検出性能を定量的に評価する。その評価指標をモデルスコアと呼び、算出式を式(1)に示す。

$$M = \frac{1}{2} \left(\frac{n_{An}}{n_{Sn}} + \frac{n_{Aa}}{n_{Sa}} \right) \dots \text{式(1)}$$

ここで n_{Sn} , n_{Sa} はそれぞれ正常状態と異常状態のサンプル数であり n_{An} , n_{Aa} はそれぞれ正常状態と異常状態において正しく状態判別できたサンプル数である。

2.3 手法の手順

本手法は2つの Step から成る。Step1 では、診断に有益な信号か否か単一な信号での調査を行なう。Step2 では、診断エージェントに用いる機械学習を用いて最良の変数組合せを予測する。

選択対象となる n 種類、 K 時間ステップ数のプロセス信号を $P_i (1 \leq i \leq n)$ とする。また、時間ステップ $k (1 \leq k \leq K)$ でのプロセス信号 P_i の値を x_{ik} 、プラントの状態を y_k とし、プロセス信号の値とプラントの状態からなるベクトル $(x_{1k}, \dots, x_{ik}, y_k)$ を時間ステップ毎に格納したデータセットを D_T とする。また、 y_k を取り除いた時のデータセット D_E とする。

Step 1 単信号で有益信号の選択

Step 1-1 モデルベーススコアの算出

D_T を学習した診断器で D_E を評価した時のモデルスコアをベースモデルスコア S_B とする。

Step 1-2 単信号のモデルスコアの算出

1 種類のプロセス信号の値 D_{Ti} を学習した機械診断器で D_{Ei} を評価した時のモデルスコア S_i を算出する。

Step 1-3 信号の選択

単信号の性能に基づいて Step2 で更に解析する信号を選択する。

式(2)の通り、ベーススコア S_B に比べて大きい S_i もしくは、モデルスコアを昇順に並べた時、順位 $\text{Order}(S_i)$ が閾値 b 以下の信号を Step2 へ用いる $S_B \leq S_i$ or $b \geq \text{Order}(S_i)$ 式(2)

Step 2 プロセス信号の組合せから最適化

Step 1 で残った m 種類、 K 時間ステップ数のプロセス信号を $P'_j (1 \leq j \leq m)$ とする。 $k (1 \leq k \leq K)$ のプロセス信号 P'_j の値を x'_{jk} とする。

Step 2-1 プロセス信号の組合せ算出

m 個の信号 P' から、1 から j 迄の組合せをまとめたデータセット $P'C_j = {}_m C_1 \cap \dots \cap {}_m C_j$ を求める。

Step 2-2 組合せからモデルスコア算出

式(3)(4)に基づいて、 $P'C_j$ からデータセット $D'E_j$ を求める。また、 $D'T_j = \{D'E_j, y_k\}$ とする。 $D'T_j$ を学習して生成した予測診断器を PLM_j とする。

$$\text{Rest}(P'C_{ji}) = \begin{cases} x'_j, & P' \in P'C_{ji} \\ \text{Empty}, & P' \notin P'C_{ji} \end{cases} \text{式(3)}$$

$$D'E_j = \{\text{Rest}(DP'C_{j1}) \cap \dots \cap \text{Rest}(P'C_{jm})\} \text{式(4)}$$

Step 2-3 モデルスコアの算出

組合せを学習した診断器のモデルスコア $P'M_j$ とする。

Step 2-4 予測診断器生成用、学習用データの生成

式(5)の通り定義したExistを用い、 $P'C_j$ の各プロセス信号 P' の存在の有無のベクトル(データセット) $D'E_j$ を求める。 $D'E_j = (Exist(P'_1), \dots, Exist(P'_m))$ のデータセットである。また、データセット $D'T_j = (D'E_j, P'M_j)$ のベクトルと定義する。-

$$Exist(P'_i) = \begin{cases} 1, & P' \in P'C_i \\ 0, & P' \notin P'C_i \end{cases} \quad \text{式(2)}$$

Step 2-5 予測診断器の作成と予測

$D'T_j$ を学習した予測診断器 PLM_j を作成する。予測診断器 PLM_j で $D'E_j$ を評価した時の予測モデルスコア PPM_j を求める。

Step 2-6 探索の終了判定

予測と実測したモデルスコアの分散や平均を算出し、そのスコア差が閾値以下である時探索を打ち切る。

そのスコア差 $Avg. S_{diff} = Average(\sum PPM_j - \sum P'M_j)$ が定めた閾値 $Avg. T$ 以下時、スコア差 $SD. S_{diff} = SD(\sum PPM_j - \sum P'M_j)$ が定めた閾値 $SD. T$ 以下時、または、 $i = N$ となった時のスコアを最終スコアとして、探索を終了する。でなければ Step 2-2 へ戻る。

3. 評価実験

もんじゅ 40%出力時のプロセスデータのシミュレーションを使って評価する。

3.1 異常の種類

- (1) 給水温度異常低下事象(Ext45)
抽気弁故障による給水加熱損失
- (2) 蒸発器熱伝達低下異常(Fouling45)
蒸発器へのファウリング付着時の異常である。
- (3) 給水系流量調節弁異常(FWCV45)
給水制御弁故障により給水流量の低下(-2.5%)した時の異常である。
- (4) 1次系主循環ポンプ流量異常(Mainpump45)
R/V(Reactor Vessel)1次系の主循環ポンプ故障によるNa流量低下異常である。

- (5) 主蒸気圧力上昇異常(SCV45)
主蒸気制御弁が故障した時の異常である。

3.2 異常検出の手順

評価用異常データから1レコードずつ運転状態を診断し、60件の診断結果の運転状態の半数以上が異常だった時をプラント異常と検出した。

最適化SVMは本手法で最適化した診断器であり、基本SVMはすべてのプロセスデータを学習して生成した診断器である。結果は次の通りである。異常発見にかかった時間が短い方の時間に下線を引いている。

3.3 異常の印加違いによる評価

3.3.1 ARモデルによるノイズ重畳

Netflow++のもんじゅ出力40%時シミュレーションデータに、ARモデルによるノイズを重畳したデータを用い評価する。ノイズは10000レコード以降に発生する。

3.3.2 結果

	評価データ	最適化SVM	基本SVM
1	Ext45	10046	<u>10040</u>
2	Fouling45	10031	10031
3	FWCV	<u>10050</u>	10582
4	Mainpump45	10173	<u>10158</u>

3.4 異常印加データを用いた評価

本評価実験では、ARモデルによるノイズを付加したシミュレーションデータに、下記プロセス箇所に異常を発生させたデータを用いて、下記評価用異常データを本最適化手法に適用する。

3.4.1 評価用異常データの詳細

■ 異常印加時間

- 連続異常：2000～4000, 5000～7000秒
- 突発異常：5000, 6000, 7000秒から20秒間

■ 印加される異常の詳細

① R/V出口Na温度：

連続異常として不感帯異常と分散増加を重畳

不感帯異常は、データ値の揺らぎが分散 1σ の範囲内の

場合、値を平均値に置換

分散増加異常は、データ値の揺らぎを3倍に増加

② 2次主循環ポンプ回転数：

連続異常として飽和異常と分散増加を重畳

飽和異常は、データ値の下側への揺らぎが分散 1σ 以上の場合、値を分散 1σ に置換

分散増加異常は、データ値の揺らぎを3倍に増加

③ 蒸発器出口蒸気温度：

連続異常：高周波減衰異常. 高周波減衰異常は、カットオフ周波数0.003Hzのローパスフィルタ

突発異常：二周期の正弦波成分を重畳. 周波数0.2Hz振幅 3σ の正弦波成分を重畳しました。(突発異常全体で4周期分)

④ 主蒸気圧力上昇異常(SCV45)：

連続異常：ピーク成分増加は周波数0.01Hz振幅 2σ の周波数成分を重畳

突発異常：二周期の正弦波成分を重畳. 周波数0.2Hz振幅 3σ の正弦波成分を重畳しました。(突発異常全体で4周期分)

3.4.2 結果

	評価データ	最適化SVM	基本SVM
1	Ext45	5263	5263
2	Fouling45	<u>5187</u>	5202
3	FWCV45	<u>5030</u>	5048
4	Mainpump45	5099	<u>2070</u>
5	SCV45	5067	<u>5060</u>

3.5 考察

本手法により異常検知速度が向上した結果があった。基本SVMに異常検知速度を劣る場合もあり、最適化できたのは一部であった。ただ、診断にはマルチエージェント診断システム上での利用を想定しており、基本SVMと最適化SVMの両方同時に検出できるため、最適化SVMで検出できなかった異常は基本SVMで検出できる。結果、どちらかいち早く検出できればよく、検知速度の向上を測ることができた。

4 結論

本研究では、マルチエージェントに特化した機械診断器最適化手法を提案した。NETFLOW++でシミュレートしたもんじゅ40%のプロセスデータに本手法を適用した。結果一部最適化する事ができた。

謝辞

本研究は、文部科学省のエネルギー対策特別会計委託事業による委託業務として、岡山大学が実施した平成21~24年度原子力システム研究開発事業「「もんじゅ」プロセスデータのハイブリッド高度処理による異常診断エージェントの研究開発」の成果である。また、日本原子力研究開発機構には、「もんじゅ」データの開示等をいただき、感謝致します。

参考文献

- [1] S. Joe Qin, Survey on data-driven industrial process monitoring and diagnosis, Annual Reviews in Control, Volume 36, Issue 2, December 2012, Pages 220-234, ISSN 1367-5788, 10.1016/j.arcontrol.2012.09.004.
- [2] H. Han, B. Gu, T. Wang, Z.R. Li, Important sensors for chiller fault detection and diagnosis (FDD) from the perspective of feature selection and machine learning, International Journal of Refrigeration, Volume 34, Issue 2, March 2011, Pages 586-599, ISSN 0140-7007, 10.1016/j.ijrefrig.2010.08.011.
- [3] Isaac Monroy, Raul Benitez, Gerard Escudero, Moisés Graells, A semi-supervised approach to fault diagnosis for chemical processes, Computers & Chemical Engineering, Volume 34, Issue 5, 10 May 2010, Pages 631-642, ISSN 0098-1354, 10.1016/j.compchemeng.2009.12.008.
- [4] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimization for support vector machines," Expert Systems with Applications, vol. 31, no. 2, pp. 231-240, Aug. 2006.
- [5] 廣安知之, 西岡雅史, 三木光範, 横内久猛. 多目的遺伝的アルゴリズムによるSVM学習データ選択手法. MPS, 数理モデル化と問題解決研究報告, Vol. 2008, No. 126, pp. 77-80, 2008.
- [6] Hiroshi FURUKAWA, Keiji KUCHIMURA, Takashi WASHIO and Masaharu KITAMURA, Information Diversification for

Intelligent Diagnosis of Nuclear Plants(in Japanese),
Journal of Nuclear Plant, Vol. 37, No. 8, 1995.

システム研究開発事業及び原子力基礎基盤戦略研究イニシ
アティブ 成果報告会資料集, 2011.

[7] 五福 明夫, 「もんじゅ」プロセスデータのハイブリッド高
度処理による異常診断エージェントの研究開発, 原子力シ

(平成 25 年#月##日)