

H-022

センサモデルと投票に基づく発電機の異常検知の検討

Basic Study on Dynamos Anomaly Detection Using Sensor Models and Voting

菊地啓¹ 近藤和樹¹ 堀田政二¹ 渋谷久恵² 前田俊二²
 Kei KIKUCHI Kazuki KONDO Seiji HOTTA Hisae SHIBUYA Shunji MAEDA

東京農工大学大学院 情報工学専攻¹Department of Computer and Information Sciences, Graduate School of Tokyo University of Agriculture and Technology
(株)日立製作所 生産技術研究所²

Production Engineering Research Laboratory, Hitachi, Ltd.

1 まえがき

産業用電力機器など社会インフラ事業における設備の安定稼働のため、機器の劣化に伴う異常予兆を速やかに検知することが求められている。予兆の検知によって劣化した機器を故障が起こる前に交換または修理することが可能となり、設備の安定性の向上につながる。本稿ではこの様な課題に対し、複数のセンサデータを用いて故障の前に機器の異常を検知するための基礎検討を行う。

設備の稼働状況を把握するために提案手法では設備内の複数のセンサに着目する。これらのセンサは冷却水の温度や油圧といった設備の稼働状況を出力する。これらのセンサを用いて異常の予測を行うにあたり重要なことは予兆が起きた際にその原因を検知結果から読み取ることである。小野田ら [1] は水力発電所の異常振動に対し、各種センサ出力と *One Class SVM* を用いた異常検出を行っている。しかし、この手法は全てのセンサ出力を一つのベクトルで表現してしまうため、異常を出力する原因となっているセンサの特定が難しい。また、センサ出力値の波形といった時系列の情報も用いていない。香川ら [2] は機器に装着したセンサ出力値の波形と *Mahalanobis Taguchi* 法を用い、機器の状況をリアルタイムで監視するシステムを提案しているが、この手法についてもセンサ出力は一つのベクトルに統合されてしまうため、どのセンサが異常の原因となっているかが不明瞭である。さらに、これらの手法は故障を検出する手法であり、予め故障を予兆するものではない。

本稿では故障の予兆とその原因を得るため手法としてセンサ毎のモデルと投票に基づく異常検知について検討する。本手法ではまず、各センサの正常時のモデルをセンサ出力の平均値を用いて作成する。次に入力データとなるセンサ出力値が正常稼働時のモデルからどれだけ外れているかを投票を用いて計算する。この投票により、ある日の各センサの出力が正常な状態と比較してどれだけ異なるかを定量的に評価することが可能となる。さらに機器の劣化に伴う故障は日々の異常状態が累積して発生する現象であると考えられるため、この投票値の日々の累積値を観察し予兆を発見する。実験では実際の発電機から得られたデータを用い、予測精度の評価を行った。

表1 各データセットの内訳

発電機名	センサ数	記録期間 [日]
A	26	519
B	21	512

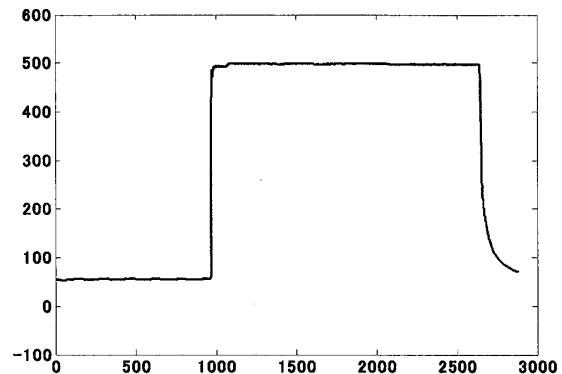


図1 あるセンサの一日の出力値の推移。

2 提案手法

以下では、複数のセンサの一日毎の平均的な出力値と実際のセンサ出力を比較することで、機器の異常を予測する手法について述べる。

2.1 曜日毎センサモデルの作成

各センサの出力値は図1に示すように一日毎に2880個の区間で記録されている。本稿で用いるデータは運転状況の異なる二台の発電機から得られたセンサ出力値を用いている。その概要を表1に示す。センサ出力値は曜日毎に平均的な形状が異なるため、日曜日から月曜日までのそれぞれの曜日に分けて2880個の各区間での平均と標準偏差を計算する。作成した平均と標準偏差を合わせて本稿ではセンサモデルと呼ぶ。あるセンサから算出した月曜日のセンサモデルを図2に示す。

2.2 センサモデルへの投票を用いた各日の異常度の算出

作成したセンサモデルを用いて各日の異常度を算出するために、入力データと入力データと同じ曜日のセンサモデルとの比較を行う。具体的には $I_k^{day}(j, t)$ を観測を始めてから第 j 日目 (day 曜日) におけるセンサ k の区間 t の出力値とし、 $M_k^{day}(t)$ を入力データと同じ曜日の区間 t におけるセンサ k の出力値の平均値、 $S_k^{day}(t)$ をそれに対応する標準偏差とする。ここで t は整数であり、 $1 \leq t \leq 2880$ の値をとる。この時、センサ k の第 j 日の異常度 $D(k, j)$

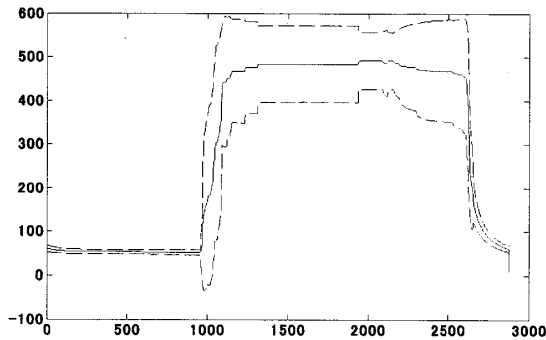


図2 月曜日のセンサモデル. 実線が平均値, 破線は平均±標準偏差を表す.

を次式で評価する.

$$D(k, j) = \sum_{t=1}^{2880} \delta(|M_k^{day}(t) - I_k^{day}(j, t)| > nS_k^{day}(t)) \quad (1)$$

ここで, $\delta(\cdot)$ は (\cdot) が真であれば1を, そうでなければ0を返す関数である. また, n は平均と平均±標準偏差との間の距離を制御するパラメータであり, 実験では $n = 2$ を用いた. $D(k, j)$ は第 j 日の入力データと入力データの曜日合ったセンサモデルとを比較し, 入力データの出力値がセンサモデルの平均± n *標準偏差外に存在した回数, すなわち異常とみなした投票回数を表している.

2.3 投票数の累積値における想定値と実測値の比較に基づく異常予測

次に, 各日の異常度を利用して異常予測を行う方法について述べる. 式(1)を用いると第 j 日にセンサ k が平均的な出力値と比較してどの程度異なった値を出力しているかを知ることが出来る. この $D(k, j)$ が著しく大きい値をとった場合は機器になんらかの異常が起きていると予想できる. この異常が累積していった場合に機器の停止等の症状が現れると考えられるため, $D(k, j)$ の日々の累積値を観察し, その累積値がある一定の値を超えたら危険を知らせる予測アラートと呼ばれる信号を出力する. ここで, 日々の機器の運転状況や気温等の変化やノイズにより, 機器に異常の無い日でも $D(k, j)$ はある程度の値をとる. そこで, 正常稼働日における $D(k, j)$ の平均値を求め, その平均値の累積値の推移と実際の累積値の推移とを比較し, 実際の累積値が平均値の累積値の th 倍を超えた時点で予測アラートを出力する. 具体的には, 正常稼働日における $D(k, j)$ の平均値を $M(k)$ とした時, 観測を始めて J 日目において次式が成立すれば予測アラートを出力する.

$$\sum_k \delta\left(\sum_{j=1}^J D(k, j) - th \times J \times M(k)\right) > 0 \geq K \quad (2)$$

ここで, K はいくつのセンサが累積値の平均値を超えたら予測アラートを出すかを定めるパラメータである. 上式が成立して予測アラートが出力された場合, 観測を初めから行う. この手法の概要を図3に示す.

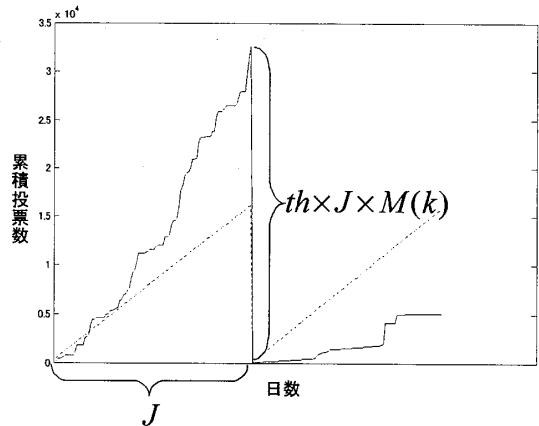


図3 累積投票数を用いた予兆検出の概要. 破線が $M(k)$ から想定される累積値, 実線が実際の累積値を示す. $M(k)$ は破線の傾きと等しい. 観測開始から J 日目において実測値が $th \times J \times M(k)$ を超えた時予測アラートを出力する. その後, 累積値をリセットして観測を再開する.

3 実験概要

提案手法がどの程度正確に異常を予測できるかを検証するため, 実際のセンサデータを用いて異常を予測する実験を行った. 実験では表1のデータセットを利用した. このデータセットにはセンサデータの他に, 実際に機器に異常が発生した場合に出力されるアラート(図6, 7上段)が含まれている.

3.1 予測アラートから実際のアラートまでの日数と予測精度の関係

提案手法の予測精度を測るため, 実験では *detect* と *purity* という2種類の尺度を定義した. これらの算出方法を示す.

$$detect = \frac{\text{実際の異常を予測できた件数}}{\text{実際に異常が発生した件数}} \times 100 \quad (3)$$

$$purity = \left(1 - \frac{\text{誤って出力した予測アラート数}}{\text{出力した予測アラートの総数}}\right) \times 100 \quad (4)$$

detect は実際の異常をどの程度予測できたかを示し, *purity* は予測アラートの精度を示している. すなわち, 両方の値が高いほど, 予測精度が高いことを示す. 本稿において, 実際のアラートが出力された日から X 日前までに予測アラートが出力された場合, 実際のアラートは正しく予測できたとみなした. また, X 日前より前に予測アラートが出力された場合, その予測アラートは誤りとした. まず, 正しく予測できたとみなす区間の日数 X と予測精度の関係を調査した. X を $X = 1$ から $X = 30$ まで変化させ, *detect* と *purity* を求めた. 予測精度の推定には, 5 fold-cross validation を利用した. すなわち, データの区間を均等に5つに分割した後, その内の1区間をテスト用のデータとし, 残りの4区間をモデル作成用のデータとして利用した. 予測アラートを出力するときのパラメータは $K = 1$, $th = 2$ とした. なお, 常に一定の

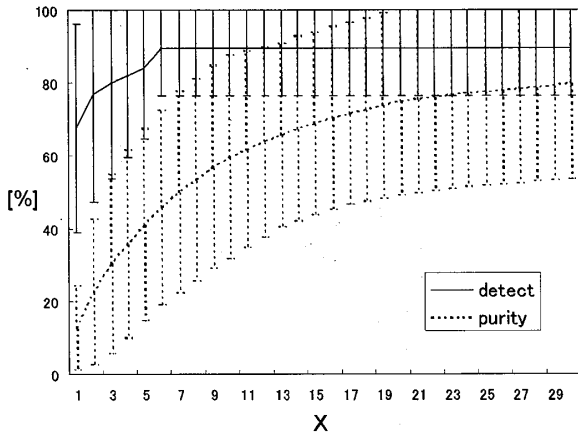


図4 AにおけるXと予測精度。

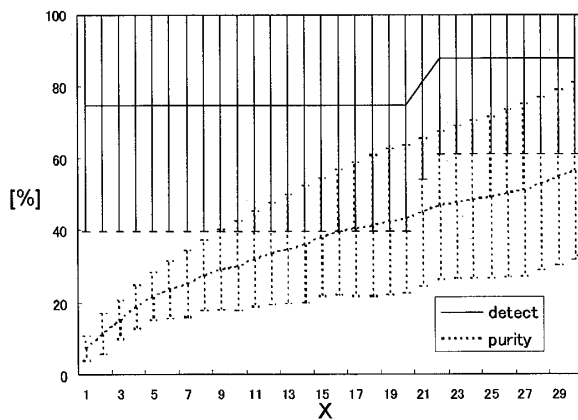
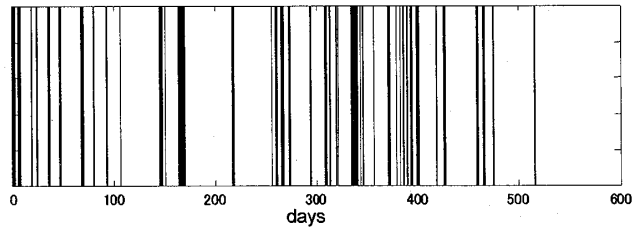


図5 BにおけるXと予測精度。

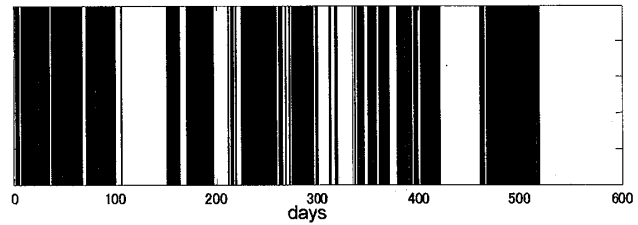
値を出力し続けているようなセンサは異常予測に適切でないと判断し、センサモデル作成の際に利用しなかった。

各データセットにおける実験結果をそれぞれ図4, 5に示す。実線が *detect*, 点線が *purity* の平均値と標準偏差を表している。図から, X が増加するにしたがって, 両方の値が高くなっていることが分かる。これは, X が増えることにより, 予測アラートを正しいと判断する区間が長くなるためである。ここで, $X = 10$ の時点において, Aでは *detect* が約90%, *purity* が約50%, Bでは *detect* が約80%, *purity* が約30%となっている。両データセットにおいて, *purity* が *detect* に比べると低い値となっている。このことから, 異常が発生する前には予測アラートを出力しているが, 正常に稼動している区間でも予測アラートを出力していることが分かる。

予測アラートの出力頻度を確認するため, 各データセットにおける実際のアラートと予測アラートの分布を図6, 7に示す。それぞれの図において, 横軸がデータの記録期間を表しており, 上段が実際のアラートの分布, 下段が予測アラートの分布を表している。図から, 実際のアラートに対して, 予測アラートが非常に多く出力されていることが分かる。

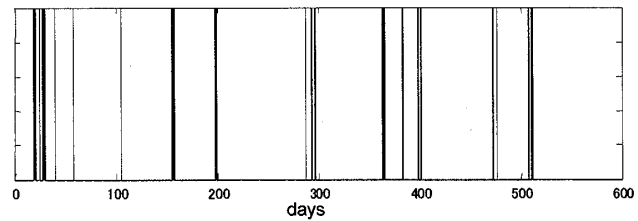


実際のアラートの分布

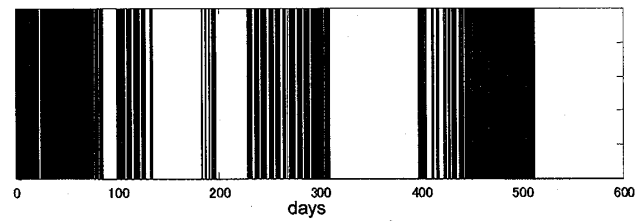


予測アラートの分布

図6 Aにおける実際のアラートと予測アラートの分布。



実際のアラートの分布



予測アラートの分布

図7 Bにおける実際のアラートと予測アラートの分布。

3.2 予測アラートの出力基準

提案手法において, 異常の予測に適切な基準を調べるため, 予測アラートの出力基準を変化させながら, 実験を行った。3.1の実験では, $K = 1$ としたが, ここでは $K = 1$ から $K =$ 全センサ数まで変化させながら予測精度を求めた。このとき, $X = 10$ と固定した。

各データセットにおける実験結果をそれぞれ図8, 9に示す。それぞれの図から, 両データセットで共通している点として, K の値が増加するにしたがって, 予測アラートの出力基準が厳しくなっていくため, *detect* が低下していることが分かる。対照的に, *purity* を見ると, Aでは $K = 17$ から増加しており, Bでは $K = 7$ から増加していることが分かる。したがって, 複数のセンサに対して, 同時に影響を及ぼす異常は少ないということが分かる。したがって, 提案手法における予測アラートの精度を改善するためには, センサ同士の相関を求め, 異常を予測するために有効なセンサの組み合わせを考える必要がある。センサ出力間の相関を用いる手法としては決定木を用いて相関の高い特徴の組を見つける手法 [3] やセン

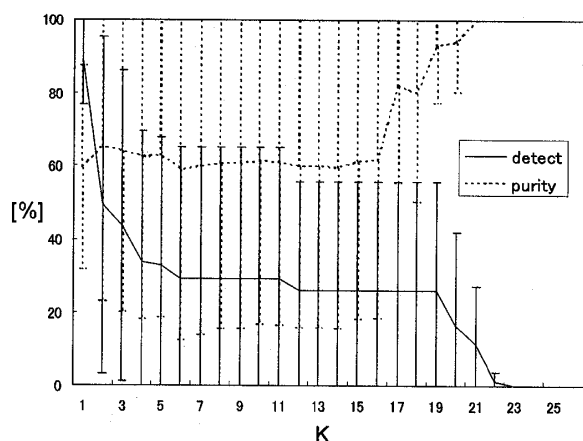


図8 AにおけるKと予測精度の関係。

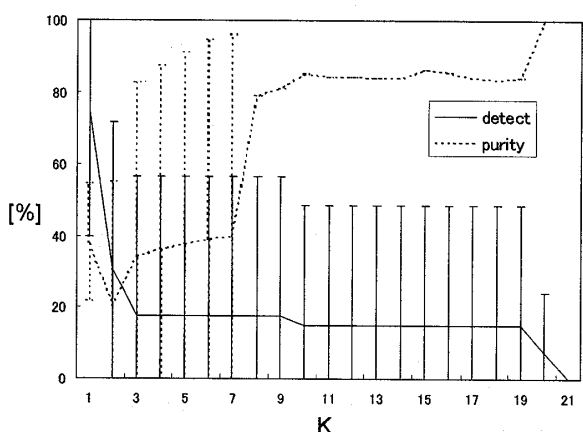


図9 BにおけるKと予測精度の関係。

サ出力間の相関そのものを用いて異常を検知する手法 [4] などが提案されているため、今後はそのような手法を取り入れる予定である。

3.3 センサ毎の予測アラート出力回数

提案手法において、各センサを単独で用いた場合に出力される予測アラートの回数にどのような違いがあるかを調べた。各データセットにおけるセンサ毎の出力回数を図 10, 11 に示す。また、パラメータは $th = 2$ とした。図の横軸はセンサ番号を表し、縦軸が予測アラートの出力回数を表している。A においては、2, 12, 25 番のセンサが、B においては 12 番のセンサが、他のセンサよりも予測アラートの出力回数が多いことが分かる。このことから、センサによって予測アラートを頻繁に出力するものとそうでないものが存在することが分かる。したがって、予測精度を向上させるために、これらセンサ毎の特性を利用するということが考えられる。なお、予測アラートの出力回数が 0 回のセンサは今回の実験において使用しなかったセンサである。

4 まとめ

本稿では、発電機といった産業用電力機器の劣化に伴う異常予兆を速やかに検知するため、複数のセンサモデ

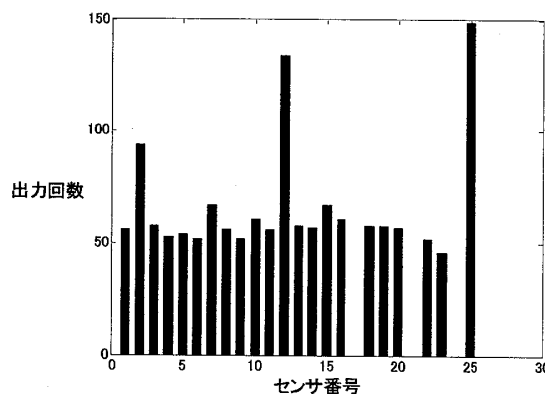


図10 Aにおけるセンサ毎の予測アラート出力回数。

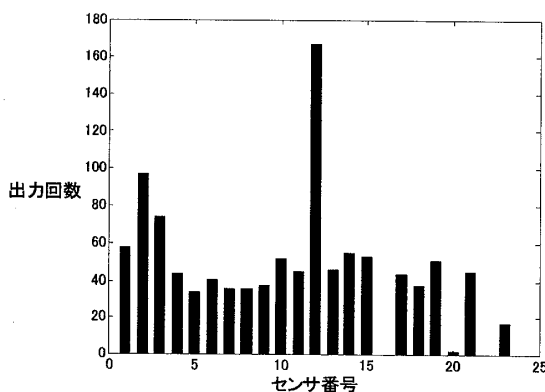


図11 Bにおけるセンサ毎の予測アラート出力回数。

ルを用いて、機器の異常を予測する手法を提案した。提案手法を用いて実験を行った結果、実際の異常が発生する前に、危険を知らせる予測アラートを出力できることを示した。しかし、正常稼動時にも予測アラートを出力し続けてしまうという問題点があることが分かった。

本稿では各センサを個別に扱って予兆の検出を行ったが、検出精度の向上のためにはセンサ間の相関の利用は不可欠だと考えられる。今後は、各センサの特性を調査し、異常の予兆検出に有効なセンサの統合手法についての研究を行う予定である。

参考文献

- [1] 小野田崇, 伊藤憲彦, 清水健二, 野邊信勝, 村上秀樹 “One Class SVM に基づく水力発電所軸受け異常振動の予兆発見 (2),” 第 19 回人工知能学会全国大会, 1F4-03, 2005.
- [2] 香川敏昌, 平井雅人, 小山茂則, 豊田佳克 “生産支援のための故障診断システムの開発,” 徳島県立工業技術センター研究報告, vol. 14, pp. 9-14, 2005.
- [3] 久保田和人, 森田千絵, 渡辺経夫 “データマイニングを用いたプラントの知的メンテナンス技術,” 東芝レビュー, vol. 61, No. 21, pp. 8-12, 2006.
- [4] 井出剛 “近傍保存原理による異常検知,” 10th Workshop on Information-Based Induction Sciences, pp. 212-217, 2007.