

# 機械学習による電気機器のエラー原因特定手法

Method to Identify Cause of Operational Errors in Electronic Products Using Machine Learning

丸地 康平 MARUCHI Kohei 曾我 幸照 SOGA Yukiteru

センサーやメモリーの低コスト化に伴い、稼働中の電気機器の状態を捉えたセンサーデータが入手しやすくなった。これらのデータには、内部機器の稼働状況や部品の状態などの情報が含まれており、うまく分析して有効活用することで、保守作業を効率化できる。家庭用燃料電池では、エラーが発生した際に、エラーに至った原因を正確かつ迅速に特定できるようになり、復旧作業を短縮できる。

東芝グループは、電気機器の保守作業の効率化を目指し、家庭用燃料電池の稼働状態を計測したセンサーデータを入力として、自動でエラー原因を特定する手法を開発した。エラー原因の診断には、機械学習手法の一つであるランダムフォレスト<sup>1)</sup>を分類モデルに適用した。実データを用いた評価実験で、診断正解率は約85%を達成し、この手法の有効性を確認した。

As an outcome of the reduced cost of sensors and memories in recent years, the collection of log data from various electronic products in operation has become easier. Effective utilization of these collected data, including the operating status of internal devices and the situation of parts, makes it possible to improve the efficiency of product maintenance work. The application of this method to residential fuel cell systems is therefore expected to reduce the time required for repair services in the event of a problem, through rapid and accurate identification of the cause of operational errors.

With the aim of facilitating the efficient maintenance of residential fuel cell systems, the Toshiba Group has developed a novel method that can automatically identify the cause of operational errors by applying the random forests machine learning method to a classification model in order to learn collected data. We have conducted evaluation experiments using actual operational data and confirmed the effectiveness of this method through the results obtained, which show that it achieves a cause identification accuracy rate of approximately 85%.

## 1. まえがき

センサーやメモリーの低コスト化に伴い、稼働中でも、電気機器の内蔵センサーのログデータが入手しやすくなっている。これらのデータには、外観からは分からない、機器内部の詳細な動作状況や部品の状態などの情報が含まれている。そのため、これらのデータをうまく分析することで、保守すべき時期の把握や部位の特定ができ、保守作業の効率化につなげることができる。

ここでは、家庭用燃料電池の保守作業を効率化するため、機械学習手法の一つであるランダムフォレストを適用したエラー原因特定手法について述べる。

## 2. 家庭用燃料電池における保守作業

家庭用燃料電池における保守作業には、定期点検のほかに、機器がエラーを発報した際の復旧作業がある。復旧作業は、顧客からコールセンターに連絡が入ることから始まる。連絡を受けたコールセンターは、保守員の手配を迅速に行う。現場に着いた保守員がまず行うのは、どうしてエラーが発生したのか、という発生原因の診断である(図1)。その際、機器の内部メモリーからエラー発報前後のセン

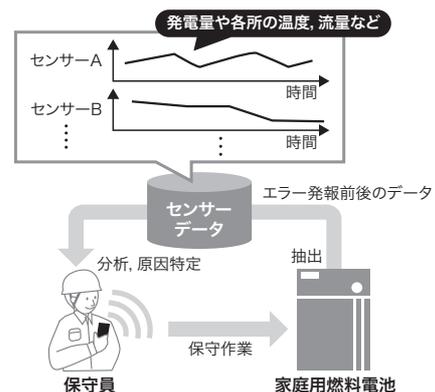


図1. エラー発報時の家庭用燃料電池における復旧作業の流れ

保守員はエラー発報前後のセンサーデータを抽出し、その挙動を可視化することで、原因の特定を行う。

Flow of repair service for residential fuel cell system in event of error

速に行う。現場に着いた保守員がまず行うのは、どうしてエラーが発生したのか、という発生原因の診断である(図1)。その際、機器の内部メモリーからエラー発報前後のセン

サーデータを抽出し、その挙動を可視化することで動作状況や内部状態を把握し、原因特定の判断材料にする。原因を特定したら、部品交換などの適切な対処を行った上で機器を再起動し、復旧を確認して作業終了となる。ここで、原因特定の誤診断は、保守作業の増加や不要な部品交換につながるため、エラー原因を正しく診断することが重要である。しかし、正確かつ迅速に診断するには豊富な経験が必要であり、そのような熟練保守員を十分に確保することは難しい。そこで、経験の浅い保守員でも正しく診断できるように、センサーデータから自動で故障原因を特定する手法が必要となる。

### 3. エラー原因の特定手法

#### 3.1 問題定義

エラー原因は、フィルター詰まりや、内部機器の動作不良、供給する空気や水の不足など多岐にわたる。エラーの発生頻度も様々で、頻度の低い原因は事例数が少なく、診断アルゴリズムを作り込むことが困難である。また、そのような原因は、正しく診断できたとしても発生頻度が低いため、保守作業の効率化への貢献度も低い。そこで、ここでは、エラー発生時のセンサーデータが十分確保できている、事例数の多い原因の特定を主目的とする。事例数の少ない原因は、まとめて“その他”というグループとして扱い、事例数の多い原因のいずれであるか、それ以外であるかを分類する問題として捉える。したがって、求められるエラー原因特定手法は、原因が分かっているセンサーデータを学習し、新たなセンサーデータの診断を行うものとなる(図2)。

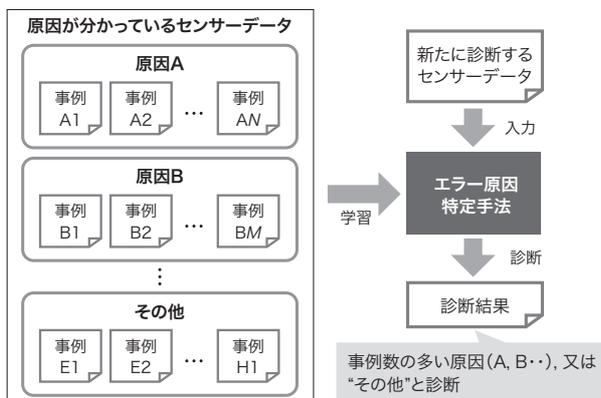


図2. エラー原因特定手法の概要

原因が分かっているセンサーデータを学習し、新たなセンサーデータを診断して原因特定を行う。事例数の少ないエラーは、その他としてまとめ、事例数の多いエラーかそれ以外であるかを分類し、事例数の多い原因の特定を主目的とする。

Overview of operational error cause identification

このように、複数センサーの時系列データから、既知の原因のどれであるかを分類する問題を、多次元時系列分類問題と言う。多次元時系列分類問題には、大きく2種類のアプローチがある。一つは、時系列データから特徴ベクトルを算出して機械学習の分類モデルに適用する、特徴ベースのアプローチである。もう一つは時系列波形間の距離を計測して距離が近いデータを同一のグループに分類する、距離ベースのアプローチである<sup>2)</sup>。3.2節で詳細を述べる今回開発した手法は、前者の特徴ベースのアプローチである。後者の距離ベースのアプローチとしては、代表的な手法に1-NN DTW (1-Nearest Neighbor with Dynamic Time Warping)<sup>3)</sup>があり、4.3節の評価実験では、この手法をベンチマークに用いる。

#### 3.2 開発した手法

開発した手法の学習時の流れを、図3に示す。まず、エラー発報があった機器の、各センサーからのデータを同じ長さの時間で区分けする。このとき、エラー発報前後でセンサーの振る舞いが変わるため、発報前〇〇秒の時間、発報後〇〇秒の時間といった具合に、発報前後で区分けするのがよい。次に、区分けした時間の各センサー値において、統計量を算出する。統計量には、平均値や、最大値、最小値のほか、傾きや標準偏差なども用いる。

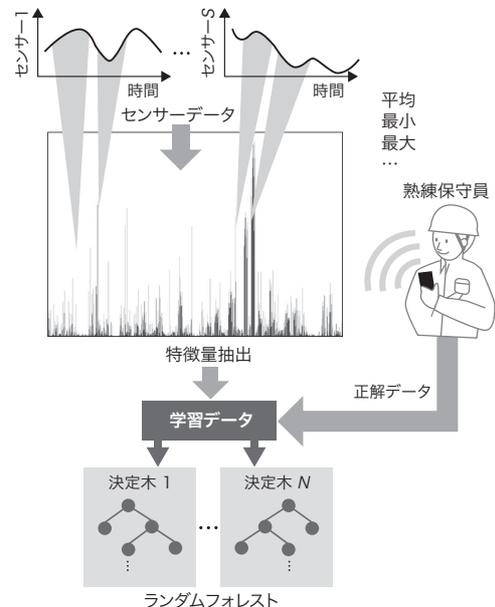


図3. ランダムフォレストを用いた学習手順

センサーデータの様々な区間から抽出した統計量の組み合わせを学習データ、熟練保守員の診断結果を正解データとし、ランダムフォレストを用いた分類モデルを構築する。

Flow of processes for operational error classification using random forests

得られた統計量の組み合わせを学習データとして、機械学習の分類モデルを構築する。分類モデルは、分類方法をアルゴリズム化して機械で行えるようにしたものであり、分類したいデータとどれに分類するかを正解データを学習することで構築できる。今回のケースでは、分類したいデータは、センサーデータから抽出した特徴量であり、正解データは、熟練保守員の診断結果を用いる。

分類モデルには、ランダムフォレストを用いる。ランダムフォレストは、決定木という古典的な木構造による分類モデルを、乱数を用いて複数個生成し、各決定木の診断結果を統合して最終的な結果を得る機械学習手法である。今回は、複数の時間区分や統計量を用いて特徴量を多数生成しているが、特徴数が多くてもうまく学習できる手法なので、ランダムフォレストを採用している。

新しいデータを診断する際は、学習して得られた分類モデルを活用し、特徴量の抽出までは学習時と同様の処理を行う。このとき、得られた特徴量を分類モデルの入力とすれば、モデルの出力が診断結果となる。

### 3.3 評価実験

評価実験のデータには、特定のエラーを発報した4種類の主要原因とそれ以外（その他）のデータを用いる（表1）。データを採取した時期には異なる2種類があり、時期Aが主に夏季に、時期Bが主に冬季に採取されたデータである。評価指標には、正しく診断した割合を意味する正解率を用いる。これらのデータを用いた評価実験の結果を表2に示す。実験ID1は、同じ時期に取得したデータで学習して診

表1. 評価実験に使用したセンサーデータの件数

Number of sensor log data used in evaluation experiments

エラー原因	時期A(夏季)のデータ件数(件)	時期B(冬季)のデータ件数(件)
A	51	10
B	31	6
C	30	4
D	30	6
その他	35	8
合計	177	34

表2. 評価実験で得られた正解率の結果

Results of evaluation experiments on accuracy rates of operational error cause identification

実験ID	実験内容	正解率(%)
ID1	時期Aのデータで学習, 評価	90.96
ID2	時期Aのデータを学習し, 時期Bで評価	73.50

ID: 識別番号

断したときの正解率を評価した。件数の多い時期Aのデータだけを用い、LOOCV (Leave One Out Cross Validation) 手法を用いた。LOOCV手法は、診断する一つのデータを除いた残りのデータを学習データとして用い、診断結果が正しいかを評価する方法であり、各データは必ず1回診断するデータとして使用される。診断するデータが学習データに含まれないので、正当な評価が行える。この結果では、正解率が90%を超え、熟練保守員と遜色のない診断正解率が確認できた<sup>(4)</sup>。一方、時期Aのデータを学習して時期Bで診断すると、正解率が約73%に低下し、学習データと診断データで季節が異なる場合に課題があることを確認した。

## 4. 季節変動への対応

### 4.1 季節変動の補正

時期Aは夏季、時期Bは冬季に取得されたデータなので、3.3節で確認された正解率の低下は、データの季節変動に起因するものと推察される。診断に有意に働いた変数を中心に調査を行ったところ、複数の温度センサーで、季節に応じてデータ分布が異なることを確認した(図4(a))。これらの温度計は、外気の影響を受けやすい位置にあるため、夏季のデータである時期Aの分布に比べ、冬季のデータである時期Bの分布が全体的に低くなった。学習したデータと診断するデータの分布が異なることが、3.3節の評価において正解率が低下した原因であると考えられる。そこで、季

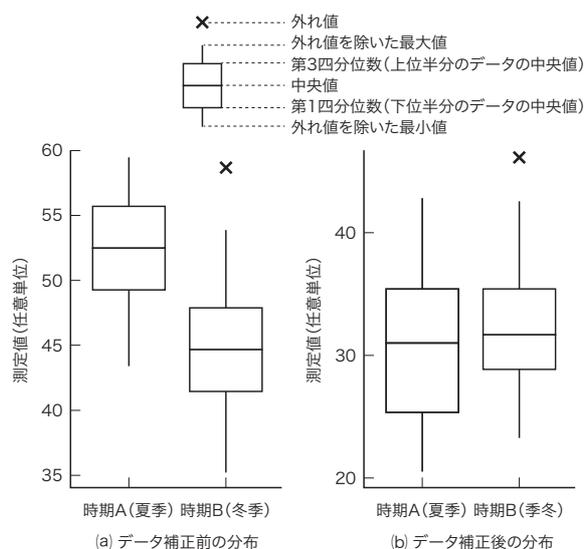


図4. 温度計データの取得時期による分布の違い

複数の温度計において、データ取得時期による分布の違いが確認されたが、データ補正を行うことで分布の違いを軽減できた。

Differences in thermometer data distribution according to seasonal variations

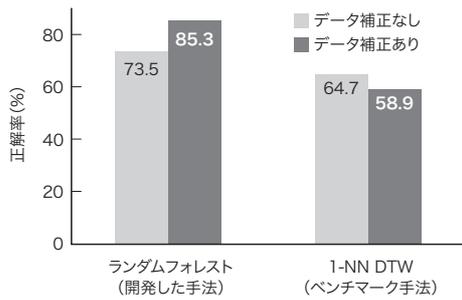


図5. データ補正の有無によるベンチマーク手法との正解率比較

データ補正により、開発した手法はベンチマーク手法より高い正解率を達成し、正解率の改善が確認できた。

Comparison of results of accuracy rate evaluations of benchmarking and newly developed methods with and without data correction for seasonal variations

節変動が生じるセンサーデータの値を補正し、分布の差を減らすことで正解率の改善を図る。補正方法としては、最も外気温に近いセンサー値との差分を取ることで行う。これにより、データ補正後の温度計の分布は図4(b)のようになり、分布の違いを軽減できた。

#### 4.2 データ補正を追加した評価

データ補正を加えたときの評価を、3.3節と同じデータで行った。すなわち、時期Aのデータを学習データとし、時期Bのデータを正しく診断できるかで評価する。また、ベンチマーク手法として、3.1節で述べた、距離ベースのアプローチの代表的な手法である1-NN DTWを用いた。この結果、開発した手法がベンチマーク手法より高い正解率を達成するとともに、データの季節性の補正により、正解率が73.5%から85.3%に改善されることを確認した<sup>(5)</sup>(図5)。

### 5. あとがき

機械学習に基づいて、センサーデータからエラー原因を自動で特定する手法を開発し、家庭用燃料電池の実データでその有効性を確認した。今後は、ほかのエラーへの展開や実際の保守作業への適用を目指していく。また、開発した手法は、特徴を抽出する区間の設定や季節性の補正の仕方が、今回の事例に特化した方法であった。ほかのドメインへの適用も視野に入れ、これらの自動化・汎用化にも取り組んでいく。

### 文献

- (1) Breiman, L. Random forests. Machine learning, 2001, 45, 1, p.5-32.
- (2) Xing, Z. et al. A brief survey on sequence classification. ACM SIGKDD Explorations. 2010, 12, 1, p.40-48.
- (3) Xi, X. et al. "Fast Time Series Classification Using Numerosity Reduction". Proceedings of the 23rd International Conference on Machine Learning (ICML'06). Pittsburg, PA, 2006-06, IMLS. 2006, p.1033-1040.
- (4) 菊池元太, ほか. "センサ情報に基づく家庭用燃料電池の故障部位判別". 2016年電子情報通信学会総合大会講演論文集, 福岡, 2016-03, 電子情報通信学会, 2016, D-8-111.
- (5) Kikuchi, G. et al. "Failure Cause Identification for Residential Fuel Cell based on Multi-Sensor Data". Proceedings of the 27th European Safety and Reliability Conference, (ESREL 2017). Portorož, Slovenia, 2017-06, ESRA. 2017, 702.



丸地 康平 MARUCHI Kohei  
 研究開発本部 研究開発センター  
 システム技術ラボラトリー  
 System Engineering Lab.



曾我 幸照 SOGA Yukiteru  
 東芝エネルギーシステムズ(株)  
 磯子燃料電池センター 燃料電池システム設計部  
 Toshiba Energy Systems & Solutions Corp.