

主成分分析を用いた設備診断

広島工業大学 ○森本直斗 ◎谷口哲至 前田俊二 青戸勇太 株式会社 明電舎 外田脩

要旨

センサデータを解析することでいち早く設備の異常を検知し、故障を未然に防ぐことが期待される。しかしデータ分析には数学の専門知識が必要なことや、数多ある分析の手法の取捨選択など実際にデータを活用するためには多くの障壁がある。そのため本研究ではそれらの障壁を減らすための分析の手法の一つを提案していく。

1章. 序章

近年、半導体の発達やインターネットの普及により多くの場面で電子機器が見られるようになってきた。それに伴い多くの企業も人件費削減のためIoTを導入し、製品の異常を人間による目視ではなくセンサを用いて判別する企業も増えてきた。総務省の調査によると、インターネットを利用している企業の割合は99.6%となっており、ほとんどの企業においてインターネットが活用されていると述べている。さらにIoTやAIなどのシステムサービスや導入状況は「導入している」が14%、「導入していないが導入予定がある」が9.8%となっており、これは昨年度と同調査よりそれぞれ2%、1.4%上昇した値である。さらに従業者規模別に見ると、「導入している」の割合は、従業者規模が大きくなるにつれて高まる傾向にあると述べている。[1][2]人の目視による検査よりセンサによる検査の方が個人差や判断ミス等が無いため、再現性が高くなり、センサ自体も年々高性能化しているため、詳細かつ大量のデータの取得が可能となっている。それらのセンサデータを解析することでいち早く異常を検知し、故障を未然に防ぐことが期待される。しかし、RやPythonに代表されるデータ分析に活用するためのプログラミング言語は広い数学の知識やプログラミングが不可欠となっており、さらに数多ある分析の手法の取捨選択やそれに紐づく基礎知識の必要性など実際にデータを活用するためには多くの障壁がある。

本研究ではそれらの障壁を減らすための分析の手法の一つを提案していく。

2章. 提案手法

本章では、提案手法の具体的な手順及び主成分分析、検定の流れについて触れていく。

2.1. 提案手法の全体像

33個のセンサデータを平均0、分散1に標準化する。その後主成分分析にかけ、第一主成分得点をグラフにプロットする。プロットした画像から予兆を判定し、それが予兆足り得るか検定を用いて判定する。上記の手順全体を図1に示す。

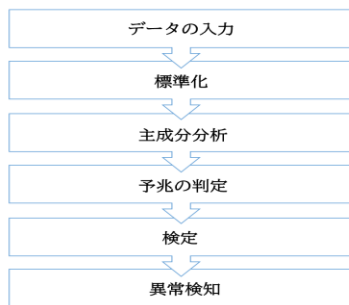


図1 提案手法の全体像

2.2. 主成分分析

主成分分析[3][4]とは、多くの量的な数をより少ない指標や合成関数に要約する手法である。主成分分析の流れは以下になる。

- Step1. 全データの重心を算出
- Step2. 重心からデータの分散が最大になる第一主成分を算出
- Step3. 第1主成分と直交方向で分散が最大となる第2主成分を算出

分を算出

Step4. 直近の主成分と直交する方向で分散が最大となる第3主成分を算出

Step5. Step4. をデータの次元分だけ繰り返す。

2.3. χ^2 検定

正規母集団 $N(\mu_1, \sigma^2)$ から大きさ n の無作為標本をとるとき、仮説 $H_0: \sigma^2 = \sigma_0^2$ (σ_0 は定数)を検定するため χ^2 分布を用いる手法である。 χ^2 検定の流れは以下になる。

Step1. H_0 が真であると仮定すると以下の統計量は自由度 $(n-1)$ の χ^2 分布に従う。

$$\chi^2 = \frac{ns^2}{\sigma_0^2} \quad (S^2 \text{は標本分散})$$

Step2. χ^2 を検定統計量として、その実現値を x とするとき対立仮説 H_1 に対応して次のような棄却域を設けて仮説の検定を行う。

両側検定の場合

$$H_1: \sigma^2 \neq \sigma_0^2$$

$$\chi^2 \leq \chi_{n-1}^2(1 - \frac{\alpha}{2}) \text{ または } \chi^2 \geq \chi_{n-1}^2(\frac{\alpha}{2}) \quad (\alpha \text{は有意水準})$$

右片側検定の場合

$$H_1: \sigma^2 > \sigma_0^2$$

$$\chi^2 \geq \chi_{n-1}^2(\alpha)$$

左片側検定の場合

$$H_1: \sigma^2 < \sigma_0^2$$

$$\chi^2 \leq \chi_{n-1}^2(1 - \alpha)$$

χ^2 検定は母集団のもつ2種類以上の特性が互いに関係があるか、あるいは独立であるかを検定する独立性の検定にも用いられ、本研究においても χ^2 検定は独立性の検定に用いている。

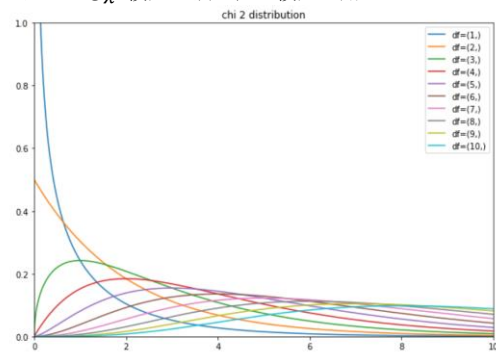


図2 χ^2 分布(自由度1~10)

3章. 研究条件及び分析結果

2章. 提案手法で述べたとおり、毎分データを取得する33個のセンサデータを標準化した後に主成分分析を行う。具体的な値は秘密保持の観点から伏せておく。ここに記載するのはセンサデータを一週間毎に分析し、結果の第一主成分得点をプロットし予兆を定義する段階からである。一度の分析に使う総データ数は、一週間の分数(10080)であり、アラートが出ている週は赤くプロットし、アラートを黒の直線で表す。アラートの前後計7日間の予兆をカウントし、アラートと予兆の関連性があるかを判断する。今回予兆は第一主成分得点の変化率が6以上のものと設定し、独立性の検定に用いるアラートの前後計7日以内に予兆有り、無し、アラート無し予兆有り、無しの4項目をカウントしていく。

続いて予兆から事前にアラートを察知するために、4項目をアラートの1日前(稼働していない日はカウントしない)の予兆の個数を用いて検定していく。

下図において、橙…予兆、赤矢印…アラートの範囲内に予兆有り、黄矢印…アラートの範囲内に予兆無し、青矢印…アラートが

無く予兆も無し、緑矢印…アラートが無いが予兆有りであり、灰線は一日の区切りを表している。

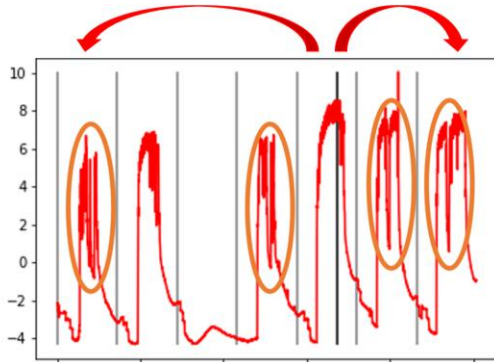


図3 予兆カウント(1)

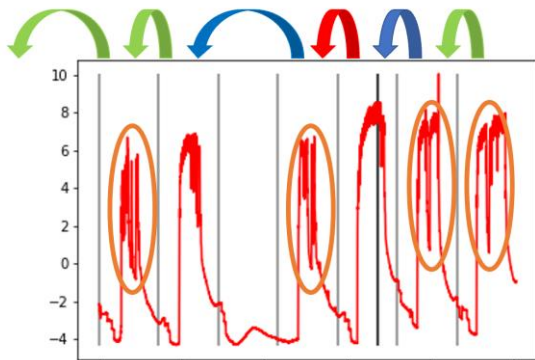


図4 予兆カウント(2)

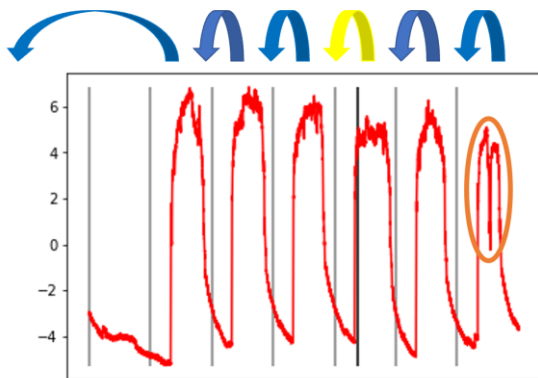


図5 予兆カウント(3)

4章. 検定による分析結果の評価

3章. 研究条件及び分析結果のカウント結果を用い、独立性の検定をする。アラートの前後計7日以内の予兆を用いて得られたカウント数で行った検定は有意水準5%では受容されたためアラートと設定した予兆の間に関連性があることが確認された。またアラートの1日前の予兆の個数を用いたカウントでは有意水準10%で受容された。

5章. 考察

ここでは3章. 研究結果条件及び分析結果および4章. 検定による分析結果の評価の結果をふまえて結果の精度向上が期待される考察をいくつか記載する。前提として統計的な分析や機械学習は扱うデータの数が多ほど結果の精度は高まるため、分析データの増量やデータを取得するセンサの高性能化が考えられるが本研究のテーマとは逸脱するためここでは掘り下げない。またここに記載する方法は一例であることにも触れておく。

本研究では33個のセンサデータを用いて分析を行ったが、前処理として扱うセンサデータを選択する方法について考えていく。分析するセンサを選択することでアラートの発報要因に即した分

析が期待される。具体的な統計的手法としてクラスタ分析[5]があげられる。一般的な階層型クラスタ分析を以下に示す。

Step1. 非類似度の定義に従って各データの最も類似度の高い組み合わせをクラスタとする。

Step2. Step1で出来たクラスタと他のデータで最も類似度の高い組み合わせをクラスタとする。

Step3. 最終的に一つのクラスタになるまで同じ手順を繰り返す。

続いて代表的な非階層型クラスタ分析の手法であるk-means法の流れを以下に示す。

Step1. クラスタの核となるサンプルをk個選ぶ。

Step2. 各サンプルを最も近い核と同じクラスタに分類する。

Step3. k個のクラスタの重心点を決めそれを新たな核とする。

Step4. 重心点の位置が変化したらStep2.へ戻り、変化しなくなるまで繰り返す。

Step5 変化がなくなれば終了する。

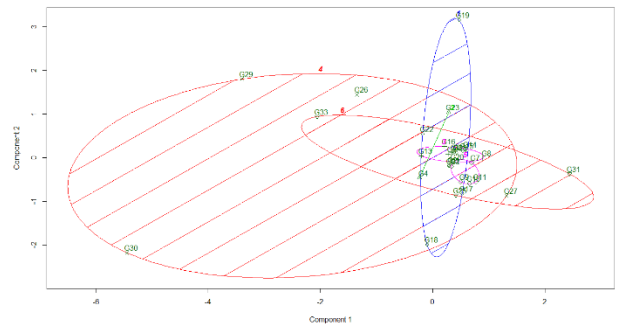


図6 k-means法結果

6章. 結論

本研究ではデータ分析を用いて設備の異常を事前に検知することを目的として主成分分析を用いて可視化する方法を提案した。そして予兆を判定し、アラートとの関連性を明らかにすることができた。

また本研究は予兆の設定方法やアラートから予兆をカウントする範囲を変更することで検定の結果も大きく変わるのでそれらの明確な設定方法を後続研究として提案しておく。

参考文献

- [1] 総務省, 平成30年 通信利用動向調査報告書(企業編)
https://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR201800_002.pdf
- [2] 総務省, 令和元年 通信利用動向調査報告書(企業編)
https://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR201900_002.pdf
- [3] 田中, 垂水 編, 統計解析ハンドブック 多変量解析, 共立出版, 1995.
- [4] Graham Upton, Ian Cool 著, 統計学辞典, 共立出版, 2010
- [5] 宮本定明 著, クラスタ分析入門, 森北出版, 1999