

データマイニングを用いたプラント異常発見

森田 千絵[†] 久保田 和人[†] 波田野 寿昭[†] 仲瀬明彦[†]

河井 研介[†] 渡辺 経夫[†] 大谷 圭子[†] 岩本 徹也[†]

株式会社 東芝[†]

1. はじめに

工場や発電所などのプラントにおいて、運転状態の異常を捉えるためにセンサの値が予め設定された一定値を超えると警報を発する仕組みが備わっている。例えば、ポンプの能力を超えた流量となった場合に警報が発せられる。一方、プラントの運転状態による適切なポンプ流量の範囲が存在する。警報が発生していなくても、この範囲から逸脱した場合、プラントにいつもと違う状況が発生している可能性が高い。そのため、プラントの運転状態に応じたセンサの適正範囲を把握し、監視することが望まれるが、つぎのような問題点がある。

- データが多種多量であるため適正範囲を手作業により定めるのは困難である。例えば、数万個のセンサの値が1分ごとに長時間にわたって記録されている。
- プラントの稼動状況によりシステムの内部状態が変化し、センサの取り得る値はそれに依存して大きく変化するためセンサの適正範囲は大域的に一定ではない。
- センサ同士の関連がある。連動するセンサやあるセンサに対して支配的なセンサなどが少なくない。

このような問題に対し、本研究ではデータマイニングによりセンサごとの適正範囲を定めることを目的とする。このうち、本稿ではセンサ間の相関を用いた分析について述べる。

2. プラントの異常検知

図1は、あるプラントにおけるあるポンプの圧力の値とプラントの生産値（例えば工場ならば製品の生産量、発電プラントならば発電量）の関係を表している。生産値が高い場合にはこの圧力と生産値はほぼ比例しているが、生産値が低い場合には1対1に対応しているとは言えない。これは、ポンプが複数存在しており、生産値が少ない場合には一部のポンプしか使われないことがある、などの理由による。そこで、シ

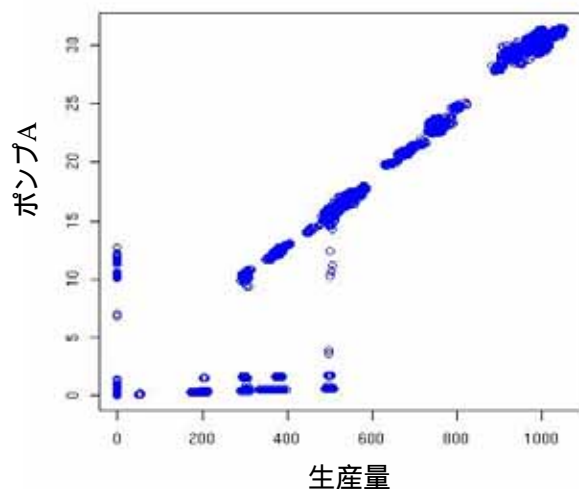


図1 ポンプAと生産量の散布図

ステムの稼動状況によってまずデータを複数の群に分類し、その後に群ごとにモデルを生成する方法が有効と考えられる。例えばこのセンサの場合には適切な分類がされれば、その時点での生産量に対するこのセンサの通常値の範囲が定められる[1]。しかし、今回のデータでは対象となるセンサの数は数千から数万にのぼり、連動するセンサも多い。そのため、稼動状況による分類やモデル生成に先立ってセンサ同士の連動の状況を調べることにする。

3. センサのグルーピング

連動するいくつかのセンサをまとめて一つのグループとすることを考える。連動するセンサをまとめたセンサ群は、他のセンサ（もしくはセンサ群）との関連を調べたり、他のセンサ（群）の値を予測したりする際に利用できる。センサの数が多い場合にはその後の分析の計算量を減らすためにも有効である。また、設計上「連動するはずのセンサ」あるいは「連動するはずがないセンサ」と照らし合わせることによる異常発見も期待できる。図2はほとんど同じ動きをする2つのセンサの例である。

Fault detection for plant system using data mining

[†]Toshiba Corporation

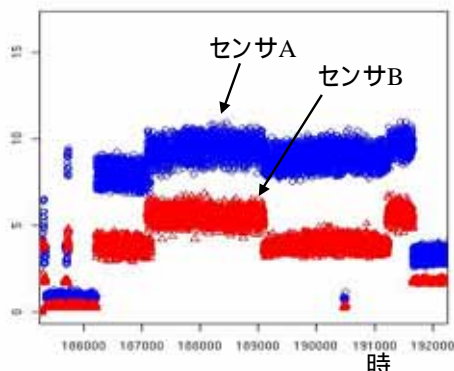


図2 連動するセンサの例

このようなセンサ群を発見するために、相関係数を類似度としてクラスタリングを行った。データ $x=(x_1, \dots, x_n)$ と $y=(y_1, \dots, y_n)$ の間の相関係数 r_{xy} はつぎの式で求められる。

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

ただし、 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 、 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ である。

センサ間あるいはセンサ群間の相関係数を求め、相関係数の高いペアからグループとしてまとめていく階層的クラスタリングを採用した。

実験では、それぞれが 40 万超の時系列データを持つ 2521 のセンサを対象とした。相関係数が 0.98 まででクラスタリングを停止場合には、2521 のセンサが 1178 のグループになった。1つのグループに含まれるセンサの数と、そのようなグループ数を表1にまとめる。

表1 相関係数 0.98 までをまとめた結果

含まれる センサ数	100-	50- 99	40- 49	30- 39	20- 29	10- 19	2-9	1
グループ数	1	2	1	2	2	22	305	842

842 個のセンサは他のセンサとグループにならずに単独のままである一方で、100 以上のセンサを含むグループもあり、連動するセンサが多いことを示している。

また、図3は、クラスタリングの集積の様子を表したものである。横軸が相関係数、縦軸がグループ数を表している。グループ数 2521 (= センサ数) からスタートし、相関係数の高い順にまとめていくにつれてグループ数が減少していく様子を見てとれる。グラフによれば相関係数が 0.9 以降でグループ数の減少が鈍り、これ以

上グループとしてまとめるのは無理があると考えられる。

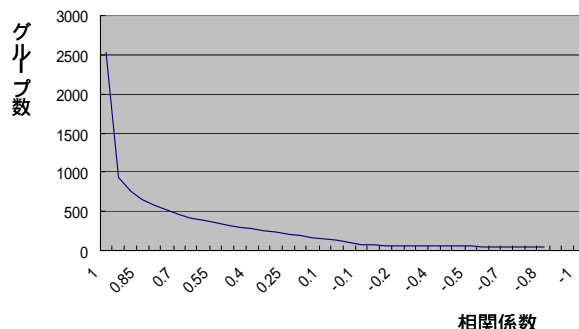


図3 クラスタリングの進行

相関係数が極端に高く連動するセンサが少なからず存在することがわかったが、相関係数をそのまま類似度として用いる際は以下の点について注意が必要である。

- 相関係数は高くても (例えば 0.95 など) 連動とは言い難いようなセンサの組もあった。データが大量であることや、相対的に大きな値に引っ張られる形で、実際には一部の値しか連動していなくても相関係数が高くなる場合があったと思われる。
- 2つのセンサの関係が深くても非線形な場合には相関係数が高くないため、関係の深いセンサのペアを見落とす恐れがある。

4. おわりに

プラントのセンサデータの適正範囲を定めることを目的としてデータマイニングの適用を試みている。本稿では特にセンサのグルーピングについて述べた。センサ間の相関係数を類似度としてクラスタリングを実施し、連動するセンサを発見することができた。この結果は、プラントの稼動状況に応じた分類やモデル生成に用いることができる。今後は、センサのグルーピングの改良と合わせてシステムの稼動状況に応じた分類とモデル生成とその自動化に取り組んでいく。

参考文献

[1]久保田、森田、波田野、仲瀬、河井、渡辺、大谷、大滝、” プラント異常発見システムにおける内部状態分類手法とその評価”, 情報処理学会第 67 回全国大会, 2005.