

放電加工機の異常検知

1. はじめに

日常の経験で、モノが壊れそうかなという勘がはたらくことがあると思います。

「洗濯機の回転音が変わる」とか「冷蔵庫がブーンと鳴り出す」とか

このような変化を定量的に（定性でもいいですが）データ化して、可視化できれば、

故障の事前対応も可能かと思いました。

たしかに、極めるには、「線形代数学」とか、「統計学」への造詣が必要で、私は不足しているのですが、

とりあえず、データを採取して検知してみて、使い物になるか否かを検討してみようと思います。

故に、2項目は飛ばしていただいて、3項目を見ていただく方が、話しが早いかと思います。

また、お願いですが、他の装置機器で、物理量（真空度、温度、電圧等）のデータを採取できないか、探しています。思い当たるものあれば、是非、一報ください。

技術部の仕事として、装置の保守管理が挙げられ、装置が必要な時に、必要なだけ利用できる状態にするのは、技術部の使命ともいえます。

また、装置には、故障という使用不能になる状態が発生し、これを無くすることは、不可能に近いです。

部品の寿命、利用者の誤使用、外乱等 回避できないな事象があります。

そこで、少なくとも、故障を予期して、早めの部品交換、装置の状態確認ができれば、

修理というプロセスは避けられないかもしれませんが、非稼働を最小限にすることは可能と考えます。

すなわち、機械学習を使って、装置の出力を、客観的な基準で変化の兆候をとらえて（＝異常検知）、対応することを、考えてみました。

また、対象する装置は、担当している放電加工機で行いました。

参考にした本は、

「入門 機械学習による異常検知—Rによる実践ガイド—」井出剛 著 コロナ社
です。Rのプログラム構文は この本を、参照しています。

2. 時系列データの異常検知

異常検知の方法として、基礎的なものとして、

- ・近傍法による異常部位検出（検証データと訓練データでその差異を評価する）
- ・特異スペクトル変異法（時系列データの変化を検知し、変化の程度を評価する）

の二つで試みてみました。

2.1 考え方

2.1.1 近傍法による異常部位検出

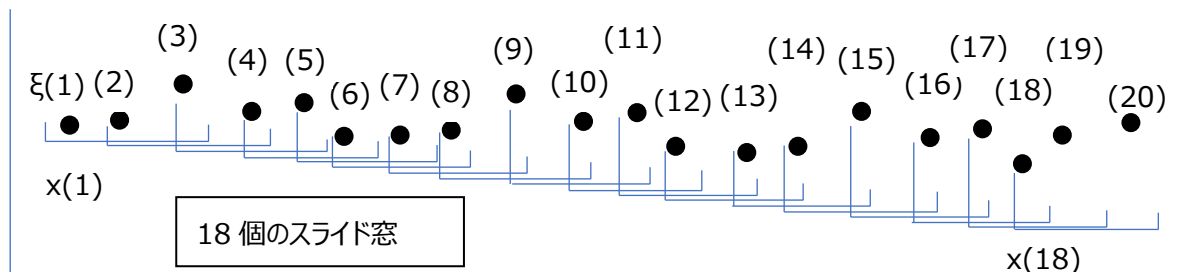
観測値として、長さ T の時系列があります。 $\xi()$ は観測値で、 T は 20 個とします。

$D = \{\xi(1), \xi(2), \xi(3), \xi(4), \dots, \xi(17), \xi(18), \xi(19), \xi(20)\}$ とします。

隣接した観測値データを w 個として、ベクトルの集まりとします。 $w = 3$ とします。

$$x(1) = \begin{Bmatrix} \xi(1) \\ \xi(2) \\ \xi(3) \end{Bmatrix} \quad x(2) = \begin{Bmatrix} \xi(2) \\ \xi(3) \\ \xi(4) \end{Bmatrix} \quad \dots \quad x(17) = \begin{Bmatrix} \xi(17) \\ \xi(18) \\ \xi(19) \end{Bmatrix} \quad x(18) = \begin{Bmatrix} \xi(18) \\ \xi(19) \\ \xi(20) \end{Bmatrix}$$

18 個の 3 次元ベクトルに変換します。 w 個の集まりを、スライド窓と言います。



時系列データを

$D = \{x(1), x(2), x(3), \dots, x(17), x(18)\}$ として、扱います。

次に、データを、訓練データと検証データを用意します。訓練データとは、異常ないと考えるデータです。

検証データとは、異常が無いかを確認するデータです。

各々の検証データについて、訓練データからどれくらいのズレがあるか距離で算出します。

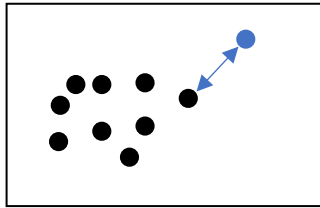
距離の算出は、R が持つパッケージ FNN の `knnx.dist` 関数 で近傍距離を算出します。

近傍距離の概略を説明します。

3次元ベクトルで進めてましたが、図にすると分かりにくいので、2次元で進めます。

黒丸が訓練データのグループです。

青丸の検証データの最も接近した黒丸との距離を近傍距離とします。



近傍距離を異常度として、時系列グラフを作成します。

閾値を超えた異常度が発生した時点を、検証することになります。

2.1.2 特異スペクトル変異法

これは、変化の度合いで、異常を判定する方法です。

観測値として、長さ T の時系列があります。 $\xi()$ は観測値で、 T は 20 個とします。

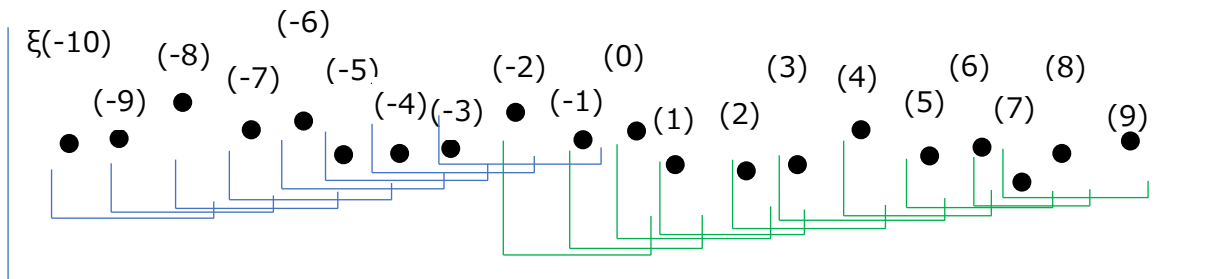
$D = \{\xi(-10), \xi(-9), \xi(-8), \xi(-7), \dots, \xi(6), \xi(7), \xi(8), \xi(9)\}$ とします。

隣接した観測値データを w 個として、ベクトルの集まりとします。 $w = 3$ とします。

$\xi(0)$ を現在の値として、 $\xi(-1)$ から $\xi(-10)$ を過去の履歴行列 X とします。

$\xi(0)$ から $\xi(9)$ をテスト行列 Z とします。

$\xi(-2)$ と $\xi(-1)$ が両行列で重なりますが、相互位置を定める整数で L (ラグ) とします。



現在時点での X と Z の行列は以下のようになります。

$$X(0) = \begin{pmatrix} \xi(-10) & \xi(-9) & \xi(-8) & \xi(-7) & \xi(-6) & \xi(-5) & \xi(-4) & \xi(-3) \\ \xi(-9) & \xi(-8) & \xi(-7) & \xi(-6) & \xi(-5) & \xi(-4) & \xi(-3) & \xi(-2) \\ \xi(-8) & \xi(-7) & \xi(-6) & \xi(-5) & \xi(-4) & \xi(-3) & \xi(-3) & \xi(-1) \end{pmatrix}$$

$$Z(0) = \begin{pmatrix} \xi(-2) & \xi(-1) & \xi(0) & \xi(1) & \xi(2) & \xi(3) & \xi(4) & \xi(5) & \xi(6) & \xi(7) \\ \xi(-1) & \xi(0) & \xi(1) & \xi(2) & \xi(3) & \xi(4) & \xi(5) & \xi(6) & \xi(7) & \xi(8) \\ \xi(0) & \xi(1) & \xi(2) & \xi(3) & \xi(4) & \xi(5) & \xi(6) & \xi(7) & \xi(8) & \xi(9) \end{pmatrix}$$

Z の X に対する変化度合いを評価するのですが、

各行列のパートンを 特異値分解 という手法でもとめます。

特異値分解の説明です。 上記の X(0)は 3 列 8 行の行列ですが、

$X(0) = U\Sigma V^*$ に分解します。 U は 3 列 3 行の行列、V* は 8 列 8 行の行列です。

Σ は対角行列で、対角成分の $\sigma_1, \dots, \sigma_q$ と並び、部分行列は、0 行列です。

また、 $\sigma_1 \geq \dots \geq \sigma_q \geq 0$ の関係にします。 σ_1 が最大値で、行列 X(0)に最も大きく影響を与えていることとなります。 σ の値を大きい方から、m 個採取して、 $U_m(0)$ の行列を求めます。

同様に、Z(0) においても $Q_m(0)$ の行列 を求めて、変化度を算出します。

(文字がダブルなので Q としました。 Z 側の U 行列とみなしてください)

次に、変化度を求めます。

「行列 2 ノルム」という量で求めます。 U と Q との空間距離を、

$\|U_m(0)*Q_m(0)\|_2^2$ で算出します。(詳細は、別途に譲ります)

変化の大きさ = $1 - \|U_m(0)*Q_m(0)\|_2^2 = 1 - (U_m(0)*Q_m(0)の最大特異点)^2$

とします。

特異値分解、変化度は、R の svd 関数で算出します。

3. 放電加工機での取り組み

3.1 メッセージデータの内容

①加工機は、自己診断機能を有しており、3段階のレベルで、それに応じたメッセージを表示します。

- ・エラーメッセージ --- 続行不可能な障害が発生した時に表示され、動作は中断します。 170 種類
- ・ハルトメッセージ --- 再開可能な障害が発生した時に表示され、動作は一時停止します。 78 種類
- ・コメントメッセージ --- 続行可能な障害が発生した時に、注意として表示されます。 58 種類

(注意というより、プロセスを表示する事が多いです)

②発生時間

[2022-06-29 20:17:43] のような 形式で表示されます。

③ その他

XYZ 座標値、プログラム名 (プログラム稼働時) 等 表示されます。

実際の 1 回分の実データは以下です。エクセルに変換できます。

153	-----									
154	メッセージ名									
155	[ERROR]									
156	メッセージNo									
157	[E00081:()OFFスイッチによる停止です。]									
158	[メッセージ発見スレッド - C2]									
159	[メッセージ発生スレッド - C0]									
160	実行モード									
161	[ラン]									
162	画面モード									
163	[ラン]									
164	メッセージ発生時間									
165	[2022-06-30 18:50:00]									
166	経過時間									
167	[00:00:04]									
168	プログラム名									
169	[FURKAN2]									
170	[FURKAN2]									
171	座標位置									
172	[G54]									
173	[X 000219.0221 Y 000017.0964 Z 000031.1697 U 000043.0689 V 000041.0000 W 000000.0000]									
174	[X 000005.1747 Y -000002.1580 Z 000000.0000 U 000000.0000 V 000000.0000 W 000000.0000]									
175	-----									

3.2 前処理されたデータ (=元データ) の内容

今回は 4月1日から6月20日の分について、データを作成しました。6317個あります。

実データから下記の元データを作成するのに、手動では面倒ですので、マクロを活用してます。

Point とは メッセージに対応する点数で、障害の程度で付与してます。

エラーメッセージ 3~4 ハルトメッセージ 2~2.5 コメントメッセージ 0.5~1.5 としました。

同じメッセージ内でも、数値を加減してます。

以下の元データを作成します。(エクセルからCSV(コンマ区切り)に変換します)

when	point
2022/04/01 08:28:01	0.5
↓	
2022/06/20 08:35:57	1
2022/06/20 08:36:10	1

元データから、異常度を評価します。

計算は R 言語 で行いました。R 言語とは、

「オープンソース・フリーソフトウェアの統計解析向けのプログラミング言語」とのことです。

Python の機械学習もありますが、

参考書が R による実践ガイド だったこともあり、R で検証してます。

また、Python も試行しましたが、R の方が異常度が顕著でした。

元データを Rstudio (R の開発環境) に取込んで、

・近傍法による異常部位検出 と ・特異スペクトル変異法 との 2 種について、

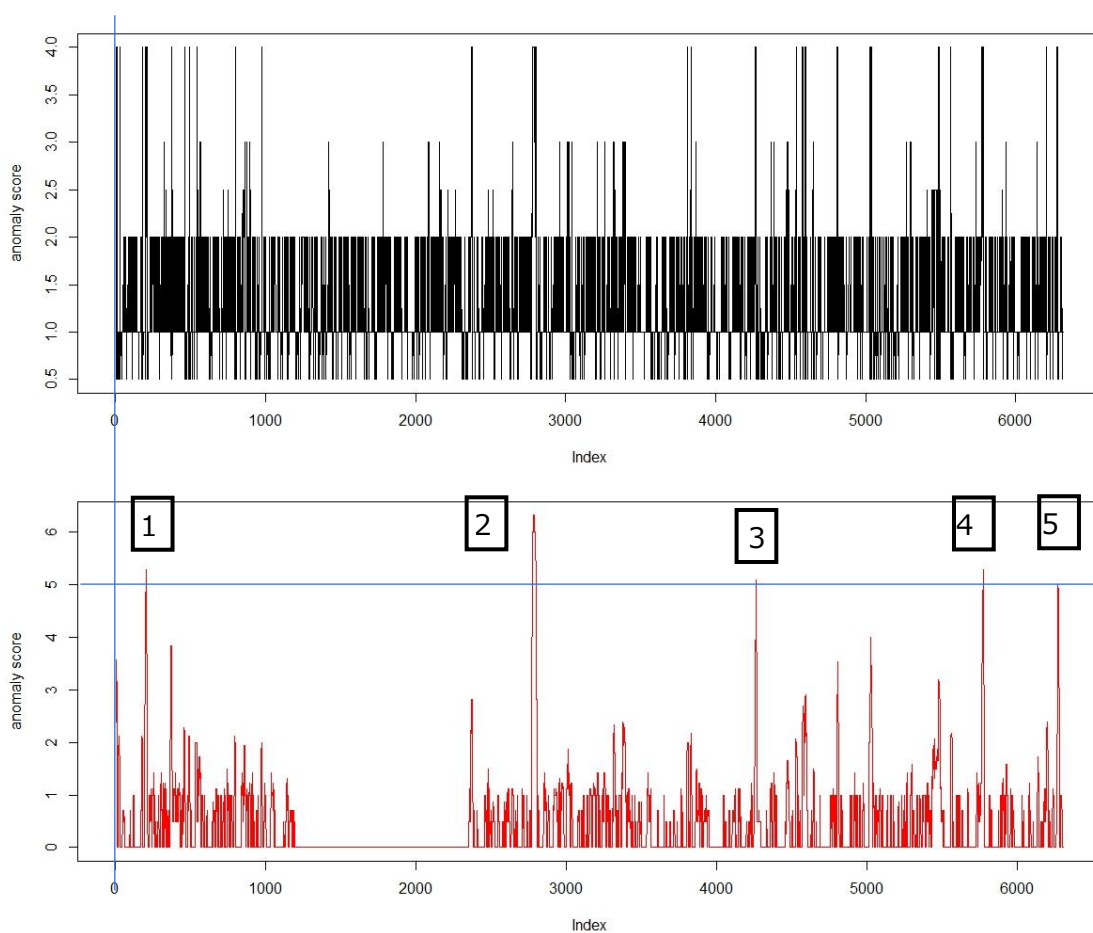
計算して、プロットしました。

3.3 結果のプロット

3.3.1 近傍法による異常部位検出（スライド窓 $w=10$ で 計算しました。）

元データは黒のグラフです。6317 個のデータに対して、0.5 から 4.0 のポイントが付与されてます。

異常度のグラフは赤のグラフです。1000 から 2000 の間で 0 になっているのは、この間は安定していたので訓練データとして採用しました。訓練データ同士を比較したので、異常度は 0 です。



異常度 5 以上に注目して、メッセージを確認しました。

①②④ は [H00207: () テンション制御装置が異常です。ENT スイッチを押して下さい。]

③ は [E00245: () 機械座標が確定していないため実行できません。]

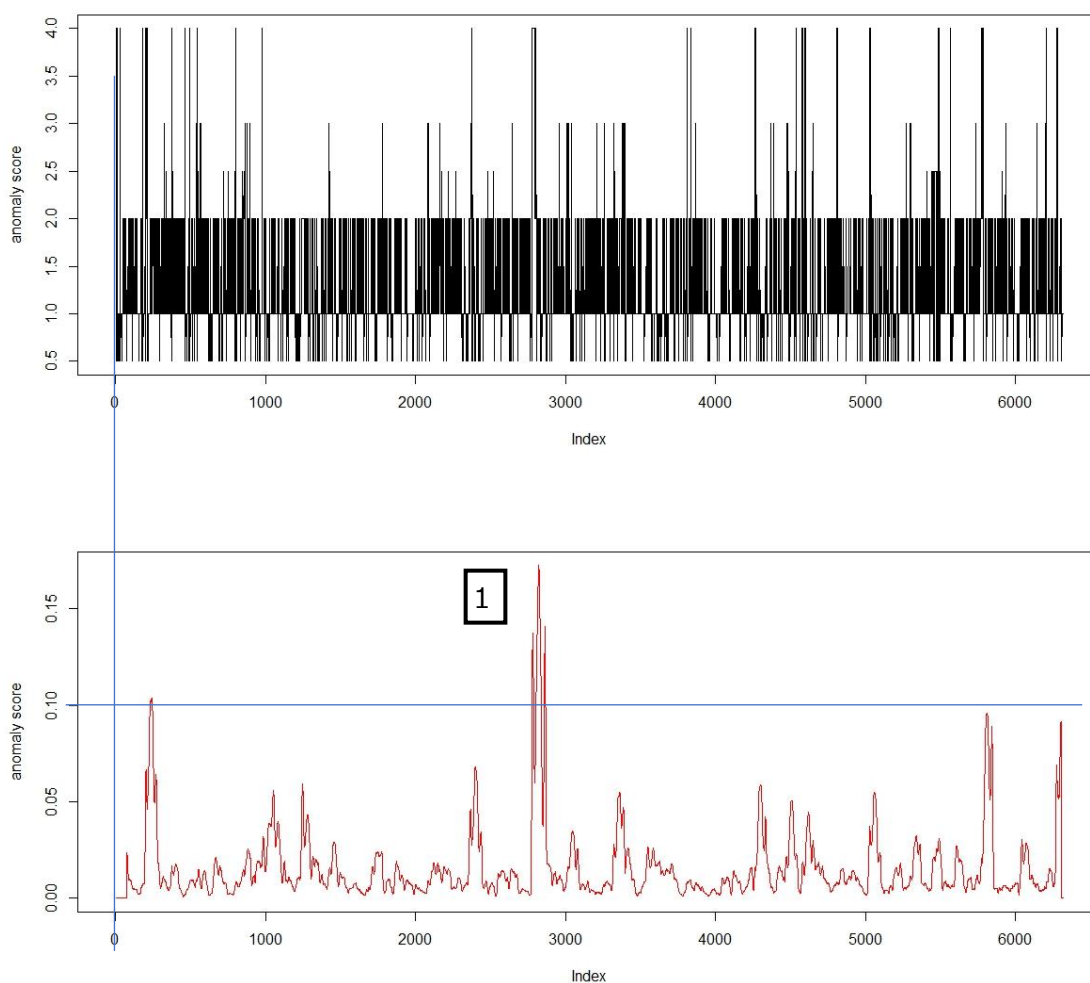
⑤ は [E00350: (V) フィードバック速度異常です。]

が 多発していました。

3.3.2 特異スペクトル変異法

($w=50$, svd 次元数 $m=1$, svd 算出用部分時系列 $k=w/2$, ラグ $L=k/2$ で計算しました。)

黒の元データは 3.3.1 と同じです。異常度のグラフは赤のグラフです



異常度 0.10 以上に注目して、メッセージを確認しました。

① は [H00207: () テンション制御装置が異常です。ENT スイッチを押して下さい。]

が多発していました。

3.3.3 プロットへの評価

元データの時点で、異常度のランク付けをしてるので、当然の結果かもしれませんが、異常度の高い時点で、不具合が発生しそうな事象が起きています。

近傍法のグラフで

①、②、④ で テンション装置が異常とあるので、予防修理が必要です。

③ の 機械座標の件は、加工機が使用不可になりました。

(コールセンターからの対応で復帰しましたが)

⑤ のフィードバック異常は、今までにないメッセージです。

元データを眺めてるだけでは、ここまでは、分かりません。

異常度グラフの方が判別しやすいのは、明白です。

3.3.4 今後の取組

継続して、データ採取して、どれくらい役に立つのか検証していきます。

時系列データの異常検知は、他の手法のあるよいですので、検討してみます。