

# クラスター分析による製品の良品自動判定と 理美容はさみの隙間形状への適用\*

二瓶 貴之\*\*

本研究では、理美容はさみの品質検査をとりあげ、デジタルデータとその分析技術を活用した製品の良否判定システムを開発した。本システムでは、理美容はさみの荷重計測装置からはさみを閉じたときの荷重変化を計測し、クラスター分析の技術を適用することで良品と不良品を分けることができる。クラスター分析はアイソレーションフォレスト法と、t 分布型確率的近傍埋め込み法 (t-Distributed Stochastic Neighbor Embedding : t-SNE) と混合ガウスモデル (Gaussian Mixture Model : GMM) を組み合わせた t-SNE+GMM 法の二つの手法を実施し、それぞれの手法の利点と欠点について検討した。アイソレーションフォレスト法では、製品の型番ごとにデータの異常検知が可能となった。t-SNE+GMM 法では、型番間で特徴を捉えたクラスター分析が可能となった。

キーワード：クラスター分析、アイソレーションフォレスト、t-SNE、理美容はさみ

## Determination of Products by Cluster Analysis, and Application to the Goodness of the Shape of Hairdressing Scissors

NIHEI Takayuki

Key words: cluster analysis, isolation forest, t-SNE, hairdressing scissors

### 1 緒言

日本では少子高齢化が進行し労働力の減少は産業界全体、特に製造業に直接的な影響を及ぼしている。この問題に対し、IoT や DX を活用した新たな解決策が提示され始めた。しかしながら IoT/DX 技術を導入しているのは主に大企業であり、中小企業では人材の不足などの要因により、技術発展の恩恵を受けられない現状がある。

理美容はさみは、熟練の職人の手によって作製されており、切れ味には、刃そのものの鋭さだけではなく、刃同士のかみ合わせも重要になるため、最終出荷時に熟練の検査官が調整を行うことで品質を維持している。しかしながら、理美容はさみの品質は、現在、世界的に統一された評価方法は確立されておらず、検査官の技能に判断が委ねられている。そのため、数値化した客観的で統一された品質評価方法の確立が課題となっている。

理美容はさみの官能評価項目の一つに「調子」があり、これは、はさみのかみ合わせを反映した使用感の評価である。この評価は、検査官の評価方法の伝承や、模倣品対策など、品質を保つための重要項目である。井上、飯村らは、独自に開発したはさみの荷重計測器 (図 1) ではさみのかみ合わせのための検査方法の標準化、規格化を図ることを目指し、研究を進めてきた<sup>1)2)</sup>。

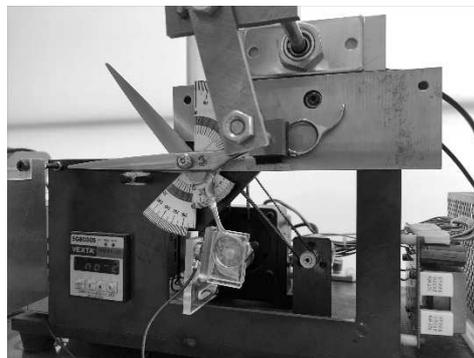


図 1 東光舎製荷重計測装置

開発したはさみの荷重計測器によって、はさみの切断現象について詳細な分析ができるようになった一方で、各々のはさみ製品の計測データを用いて、職人が実施している調子の良否の判断を行えるまでには至っていない。

そこで本研究では、県内企業で製造されている理美容はさみ製品で実際に行っている検査工程を例とし、デジタルデータとその分析技術を活用した良品自動判定システムの開発を検討した。本システムを適用することで、熟練者の技能が必要な検査工程を、誰でも実施できるよ

\* 令和 5 年度ものづくり DX システム導入支援強化事業

\*\* 電子情報システム部

うになることが期待される。

本システムでは、データサイエンスの分野で近年急速に発展している技術の一つである、クラスター分析を用いた。計測装置から取得される製品の状態を表すデータから、異常な計測データを自動で判定する方法を検討した。

## 2 実験方法

### 2-1 理美容はさみ用荷重計測装置による評価

図2は東光舎製のはさみ開閉荷重計測装置のモデル図である。当該装置ははさみを固定するクランプ機構、はさみを開閉させる駆動部からなる。駆動部はステッピングモータとモータに接続しはさみの持手部分と接触する回転アーム、アーム先端のはさみと接する部分に取り付けられた閉口時にかかる荷重を測定するロードセル、及びロータリーエンコーダで構成される。

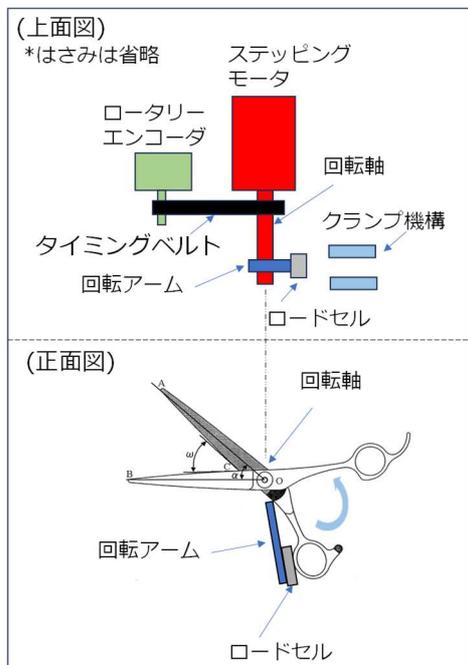


図2 はさみ開閉荷重計測装置モデル図

本装置では、はさみを開口状態からステッピングモータにて一定の角速度で閉口させ、その時にはさみの持ち手にかかる荷重の変化(波形と呼ぶ)を計測する。時間が経過するにつれて、はさみはアームの回転力によって徐々に閉じられていく。この時の荷重がロードセルによって計測され、データロガーに記録される。本装置では毛髪や試験用紙など被切断物の切断の際の荷重波形を計測することができるが、本実験においては、被切断物を用いず何も無い状態での荷重波形を計測する。これを空切り波形と呼ぶ。

図3に検査官により良品と判定された理美容はさみの空切り波形の例を示す。荷重が最大になる時が、交点が刃

先に到達した時である。閉口時の荷重波形において、開始後から荷重が最大になるまでの領域が、製品の「調子」を計測するための、かみ合わせを最も反映している箇所である。

本実験は、東光舎製品6種129丁の理美容はさみと、理美容はさみの品質規格の制定を目指す団体 Hair-dressing Scissors Consortium (以下、HSC)の各社より理美容はさみを合わせて9丁入手し、合計138本の荷重計測データを収集して分析の母集団データとした。

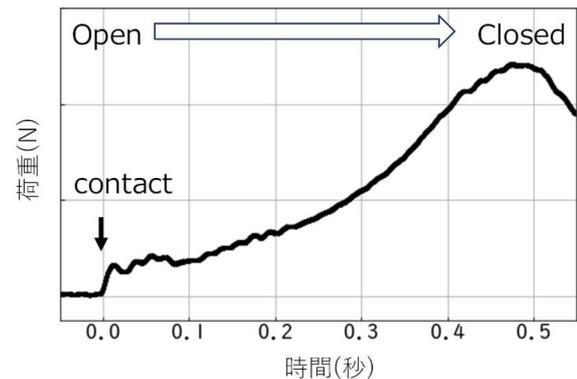


図3 はさみ(良品)の荷重計測例

### 2-2 クラスタ分析

クラスタ分析は、多数のデータについてそれぞれ特徴が異なるデータを集団(クラスタ)として分類するためのデータ分析手法であり、教師なし機械学習の一種である。すなわち、ヒトの手によって、「正常」「異常」といったラベルづけをすることなく、データを分類するために使われる手法である。

本研究で検討したアイソレーションフォレストは、クラスタ分析の一手法であり、階層構造で、多数データと小数データに区別する異常検知手法である<sup>3)</sup>。また、t分布型確率的近傍埋め込み(t-Distributed Stochastic Neighbor Embedding、以下 t-SNE)と混合ガウスモデル(Gaussian Mixture Model、以下 GMM)を組み合わせたクラスタ分析も検討した。

一つ目のアイソレーションフォレスト法は、決定木を用いて異常検知を行うため、荷重計測データのような高次元のデータに対して効率的に検出できる特長がある。また、木構造における末端ノードまでのパス長の長さで異常度を計算するため、他の一般的な異常検知手法よりも負荷を少なくして計算できる。

二つ目の手法である t-SNE 分析は次元削減手法の一種であり<sup>4)</sup>、また、GMMはEMアルゴリズム<sup>5)</sup>を用いて最適化を行う非階層クラスタ分析の一種である。t-SNEとGMMを組み合わせて計測データに適用することで、はさみの型番ごとの波形の特徴を複数のクラスタとして認識することができるため、どの型番にも属さない計測

データを異常として検出できる。したがって、複数のはさみの型番を同時に取り扱うことができる。例えば、特定の型番で計測データが一つしかないような場合でも異常判定できるため、検討手法の一つとして採用した。

以上、二つのクラスター分析手法を、理美容はさみの良品/不良品を自動的に判定するためのツールとして検討した。

データ処理は Python を用いて行い、各分析は Python の機械学習ライブラリ `scikit-learn` で提供されている `sklearn.ensemble.IsolationForest`、`sklearn.manifold.TSNE` 及び `sklearn.mixture.GMM` クラスモジュールを用いて分析を実施した。可視化には同じく Python のグラフ描写ライブラリ `Matplotlib` を用いた。

### 3 結果及び考察

#### 3-1 アイソレーションフォレスト法と分析結果

東光舎の理美容はさみ製品 6 種 129 丁について、型番ごとに閉口時の荷重計測を実施し、それぞれの結果に対してアイソレーションフォレスト法による分析を実施し

た。

例として、型番 A 及び B のアイソレーションフォレスト分析による良品自動判定の結果を図 4 及び図 5 に示す。各図の左側のグラフは個々の荷重計測の波形データであり、赤い波形が自動的に「異常」の予測ラベルが付されたデータである。

また、各図の右側のグラフは、各結果に対してアイソレーションフォレストによって評価された異常度  $s$  (Degree of Anomaly) のスウォームプロットである。異常度  $s$  は原論文<sup>3)</sup>によれば、以下の式で与えられる。

$$s = 2 \frac{E(h(x))}{c(n)}$$

ここで、 $x$  は決定木上にある点、 $h(x)$  は点  $x$  のパス長 (ルートから点  $x$  までの長さもしくは深さ)、 $E(h(x))$  は全ての点  $x$  で計算されるパス長の期待値 (平均値)、 $c(n)$  は二分木の構築に失敗したときのパス長である。上記異常度  $s$  は 0 から 1 まで変動し、 $s$  が 1 に近づくとき異常、 $s$  が 0 に近い場合は正常となる。これらは相対的であり、すべての点で  $s=0.5$  となる場合、分析したデータセットのすべての

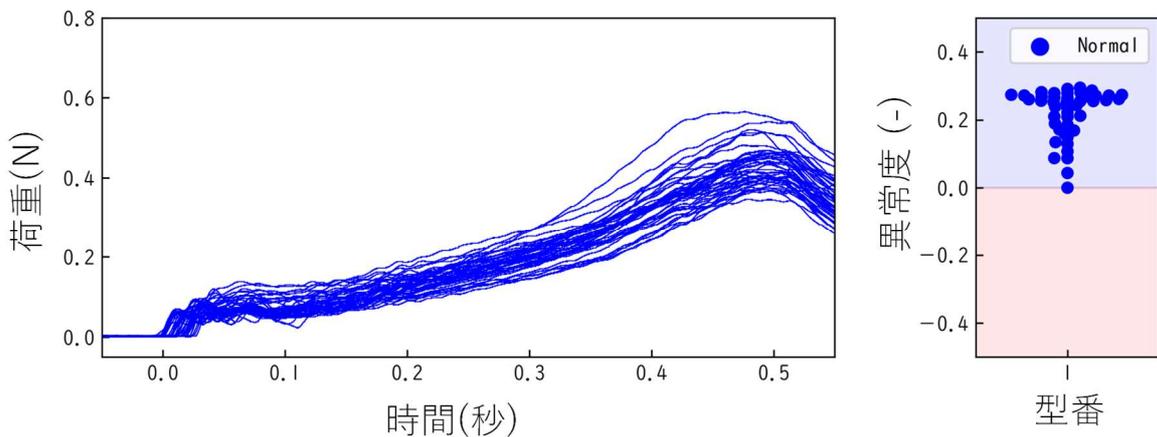


図4 型番 A のアイソレーションフォレスト分析による良品自動判定

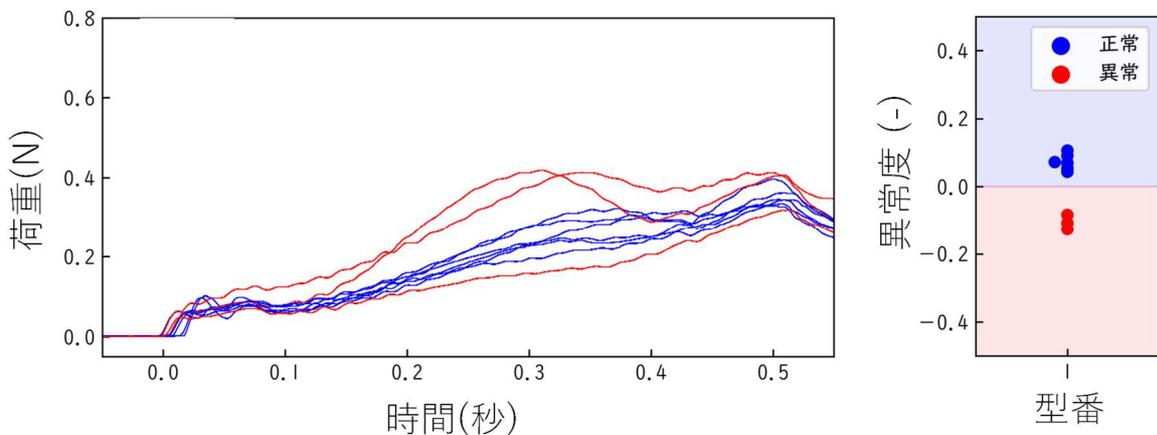


図5 型番 B のアイソレーションフォレスト分析による良品自動判定

データに明確な異常がみられないことを意味する。

Scikit-learnにおけるsklearn.ensemble.Isolation-Forestでは、-1から1まで変動し、異常度sが負の値になっているものが、「異常」と判定される。

なお、スウォームプロットとはデータ点の分布と値を同時に表現する可視化手法であり、データ点が重なる場合には横に広げて表示している。

理美容はさみ製品のそれぞれの型番について、荷重曲線とアイソレーションフォレストの予測、異常度の分布を比較しながら予測結果について考察した。出荷数量が多い型番Aでは、全ての荷重計測波形が安定しておりバラツキが少ないことが分かった。このため、図4に示すように、型番Aの異常度のスウォームプロット(図の右側)では全てのデータが近くに密集しており、「異常」を示す赤い波形が無いことがわかった。これは、全てのデータに明確な異常がみられないことを意味する。

これに対して、図5に示す、出荷数が限定的な型番Bのスウォームプロットを見ると、異常データは正常データと比べて明確に異常度(赤い点)が分離されており、それらに対応した荷重計測波形(赤い波形)も各型番を代表するデータ(波形)と比べてその形状が異なっていることがわかる。

このように、アイソレーションフォレストは10丁程度の計測データでも、異常を的確に自動判定できることが分かった。

しかし、アイソレーションフォレストによる異常検出は高精度である一方で、同一の型番のはさみで複数の計測データが必要である。はさみの品質評価に適用する際に、依頼された特定の形状(型番)のはさみを一つだけ評価したい場合がある。そのような場合にはリファレンスとなるデータがなくアイソレーションフォレストによる異常検知は難しい。そのため、t-SNEとGMMによる異常検知手法について検討した。

### 3-2 t-SNEとGMMによる異常検知とクラスター分析結果

図6は、型番A、型番B、型番Cの荷重計測波形を同じグラフに図示したものであり、それぞれ型番ごとに異なる形状(刃の長さ、太さ等)に対応した波形である。これらをt-SNEによって次元削減し、1軸目と2軸目の二次元散布図にプロットしたものが図7である。t-SNE変換後においてもそれらの違いを反映した分布を持っていることがわかる。

t-SNEにより変換された二次元座標は相対的であり、分析するデータの持つ特徴に依存し、物理的な意味を持たない。しかし、各データ点の相対位置の座標は、型番ごとに刃の長さや隙間形状が異なることによる荷重計測波形の平均的な傾きの違いや、はさみが閉じ切った際の最大荷重などの違いとして現れる。

図6における赤矢印で示した型番Bの明らかに異常な波形2データは、t-SNE変換後の図7の散布図において赤矢印で示した2点としてプロットされ、他の波形計測データのt-SNE座標から大きく外れており、明らかに異常値として判定できる。

さらに、これらをGMMでクラスター分析することで、その点がどのデータ群に属するかを判定し、その座標に点が生成される確率(もしくは尤度)を計算することができる。

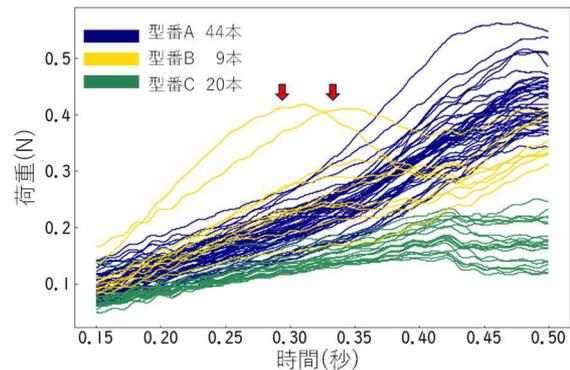


図6 型番A, B, Cの荷重計測例

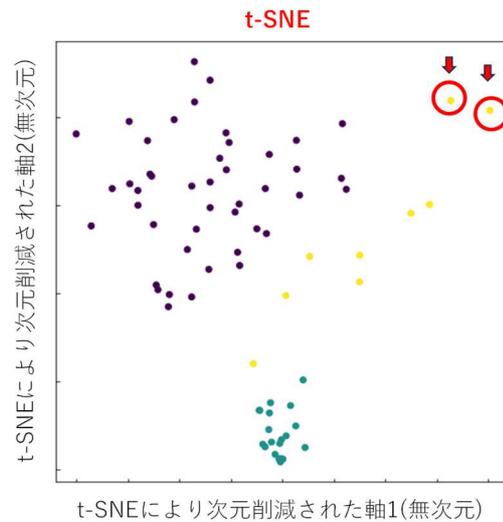


図7 型番A, B, Cのt-SNEによる次元削減

図8はt-SNEの結果に対して、GMM クラスター分析を実施し、各クラスターから生成される確率に比例する尤度を等高線として表したグラフで、等高線の中央付近ほど生成確率が高く正常なデータであることを表している。点がある程度まとまった集合が正常なデータであり、そこから外れた点は異常データと判断することができる。図8から明らかなように、赤矢印で示した点は尤度が低く、したがって、異常なデータ点であると推定される。

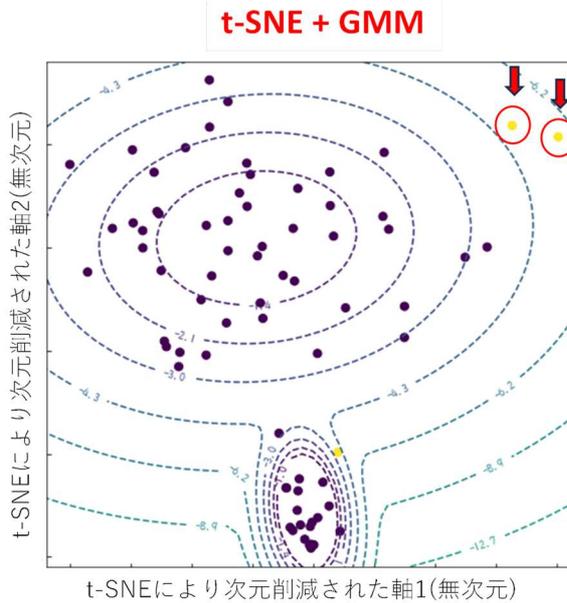


図8 t-SNE+GMM クラスター分析結果  
等高線は確率に比例する負の対数尤度  
(Negative log-Likelihood) を表す。

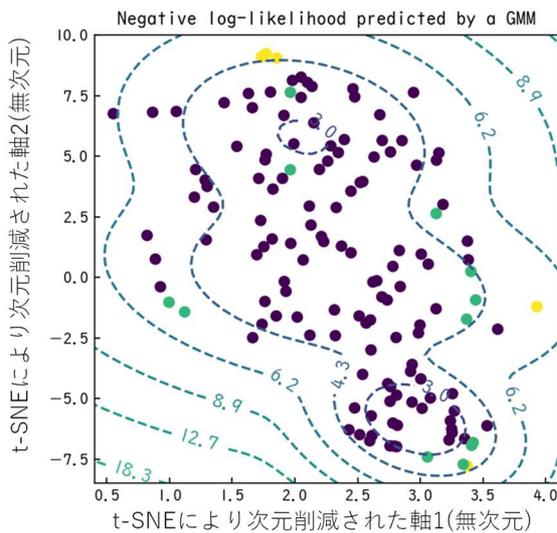


図9 前処理済みデータのt-SNE+GMM  
クラスター分析結果  
東光舎製6種129丁(青色,黄色)及び  
HSC団体5社9種9丁(緑色)の理美容  
はさみを示す。

また、図8からは、t-SNE変換が元のデータである荷重計測の波形の形状に大きく依存することが分かる。これが影響し、特定の型番で計測データが一つしかない場合について判断をしたいときに、他の形状の波形データを参考にして判断することができない。

そのためデータ前処理を実施し、他の型番でもt-SNE変換後に同じ分布を持つようにする必要がある。具体的には、計測データに微分と移動平均およびローパスフィルタを実施する。それによって、型番ごとの右肩上がりの傾きを平滑化する。この判定精度について検討した。

図9は前述のように前処理したデータについて、t-SNE分析およびGMMクラスター分析を実施した結果である。図中各点の色は、アイソレーションフォレストによってそれぞれ色分けしたもので、正常ラベルが付された荷重計測波形のデータ(青色)、異常ラベルが付された荷重計測波形のデータ(黄色)及び、HSC団体の型番が異なる9丁の荷重計測波形のデータ(緑色)である。

前処理後でも、異常ラベルが付されたデータは負の対数尤度が高く算出されているが、正常ラベルが付されたデータも同程度に負の対数尤度が高い値で算出されている点もある。

このことから、現在の前処理条件は、ある程度の正常/異常判定が行えるが、その判定の精度の向上が課題として明らかになった。判定精度の向上のためには、サンプル数や種類の増加、勾配量や移動平均など、前処理の条件について更なる検討が必要である。

#### 4 結言

本研究では、データサイエンスによるDX化技術を地域企業に技術展開するために理美容はさみを取り上げ、アイソレーションフォレスト及びt-SNE+GMMによるクラスター分析を用いて、理美容はさみ製品の良品/不良品判定について検討した。

得られた結果は以下のとおりである。

- (1) アイソレーションフォレスト法では、型番ごとに波形の異常検出が可能となった。しかし、測定データが一つしかない型番に対しては、リファレンスとなるものがないため検出できなかった。
- (2) t-SNEでは、型番ごとの波形を二次元平面図にプロットすることで可視化でき、その結果をGMMクラスター分析することによって異常/正常の判定ができるようになった。しかしながら、型番ごとにデータが一つしかないときに判定したい場合にはデータの前処理が必要であり、適切な前処理の条件を得るための更なる検討が必要である。

#### 謝辞

本研究は株式会社東光舎をはじめとするHSCの各社の御支援により、成果を上げることができました。特に、株

式会社東光舎の井上代表取締役社長には多大なる御協力をいただきました。心から感謝の意を申し上げます。

#### 文 献

- 1) 井上研司：理美容はさみの切断特性と切れ味の定量評価に関する研究, 早稲田大学博士学位論文(2006)
- 2) 飯村崇：精密はさみの高付加価値化技術に関する研究, 岩手大学博士学位論文(2014)
- 3) Liu, F. T., Ting, K. M., Zhou, Z. H.: Isolation forest, Proceedings of the 8th IEEE International Conference on Data Mining (ICDM'08), Pisa, Italy, p.413-422(2008)
- 4) van der Maaten, L. J. P., Hinton, G. E.: Visualizing Data using t-SNE, Journal of Machine Learning Research, 9, p.2579-2605(2008)
- 5) Dempster, A. P., Laird, N. M., D. B. Rubin: Maximum-Likelihood from Incomplete Data via the EM Algorithm, Journal of Royal Statistical Society, Series B, 39(1), p.1-38(1977)