

生産工程における不良要因検知と エミュレーターの推定[†]

大坪洋介, 大谷直也, 近未恵美, 杉山 将

Defect Factor Detection in Production Processes and Emulator Estimation

Yosuke OTSUBO, Naoya OTANI, Megumi CHIKASUE and Masashi SUGIYAMA

精密機器は大量の部品や部組から構成されている。実際の量産工程においては多数の製品を管理しているため、検査工程に対応する下流のデータは取得できるが、上流にある部品パラメーターを取得することはできない。したがって、工程中に仕様を満たさない不良品が発生した場合、人的に不良要因を調査し特定する必要がある。本研究では、近似ベイズ推定に基づいた生産工程における不良要因検知手法を提案する。この手法は、次の2ステップで構成される。1) 生産工程シミュレーターを活用し、データから逆問題を解く、2) 不良が発生したデータと正常データを用いて、各部品パラメーターに対する不良要因スコアを計算する。実際のシミュレーターを用いて検証し、合理的な結果が得られることを示す。さらに、スパースモデリングを組み合わせることで、シミュレーターを模倣する疑似的な数理モデルを構成できることを示す。

Precision mechanical products consist of lots of parts and assemblies. Data accumulation systems are usually installed to monitor production processes, but only downstream data can be acquired with respect to the processes of precision equipment. So, when a defect occurs, mechanical experts must manually analyze a physical product to identify the cause (factor). In this study, we propose a novel detection method for the defect factor based on the approximate Bayesian computation (ABC) with a design simulator. Our method consists of two stages: 1) solve the inverse problem from the data; 2) derive defect factor scores from two datasets corresponding to normal data and data with defects. Numerical experiments and an application to actual data yield consistent results with design information given as domain knowledge. Furthermore, it is shown that a surrogate model, i.e., an emulator, that imitates a simulation can be estimated by combining a sparse modeling.

Key words 生産工程, 不良要因検知, エミュレーター推定, 近似ベイズ計算, スパースモデリング
production processes, defect factor detection, emulator, approximate Bayesian computation, sparse modeling

1 はじめに

近年、生産工程に設置したセンサーをネットワークでつないで構築したフィジカル空間のデータを、データ分析技術を使ってサイバー空間で分析・モニタリングし、分析結果をフィードバックして生産工程を高度化するというサイバーフィジカルシステムが注目されている [1]。精密機器の生産工程においても、少量多品種生産が主流となり開発期間が短縮されたため、工程や測定機などのデータを収集・分析し、生産工程の自律化や不良要因特定などが望まれている。特に、加工装置データや中間の製品出来栄データなどを使い、測定されていない部品寸法、装置の設定などを予測する技術は仮想計測 (Virtual Metrology) と呼ばれ、主に半導体製造工程で活用されている [2], [3]。

仮想計測は量産工程において不良要因等を特定する際に有効であるが、予測したいデータの項目数に対し測定データの項目数が少ない場合には、そのまま適用することができない。

精密機器の生産工程は、大量の部品を部組から総組状態まで組立てる組立工程と、部品・部組の位置や姿勢などを調整する調整工程、製品が最終性能を満たすことを検査する検査工程など多数の工程から構成され非常に複雑である。部品寸法は最終性能に大きな影響を与えるが、厳しい公差を設定しすぎると低コストでの量産が難しくなる。そこで、光学製品の設計では機械部品と光学部品に適切に公差を配分できるよう、設計段階で量産時の光学性能・量産性・公差感度を評価可能な生産シミュレーター [4] などが利用される。量産中も部品寸法をモニタリングできれば不良要因

[†] 本稿は、引用文献 [11] に対してエミュレーター推定の考察を加えたものである。

特定に活用できるが、部品数が多いためすべての部品を測定することはコストの面で現実的ではなく、抜取検査などによって品質を確認している。

本研究の第一の目的は、生産工程から取得される限られたデータから、不良要因となる部品パラメータを検知することである。ここでの不良とは、量産された製品群の性能が設計仕様を満たさないことを指す。具体的には、設計仕様を満たす正常のデータセットと不良品を含むデータセットが与えられた際に、両者の差分から不良要因を検知することを目指す。一般に、取得できるデータは検査工程を経た総組の性能データであり、大部分の部品パラメータのデータは取得できないため、データ駆動型の解析だけでは不良要因は検知できない。そこで、部品パラメータから最終性能が出力される生産シミュレーターと生産工程データを組み合わせることで不良要因検知を実現する。また、本研究の第二の目的は、エミュレーター（疑似モデル）を構成することである。エミュレーターとは、一般に複雑なシミュレーターを模倣する疑似的な数理モデルである。本研究では、スパースモデリング [5] によってシミュレーターと同等の結果を与えるエミュレーターを構成することができ、最終性能に寄与する変数を自動で選択できることを示す。

2 問題設定とアプローチ

Fig. 1 は生産される製品の生産工程の概略を示している。

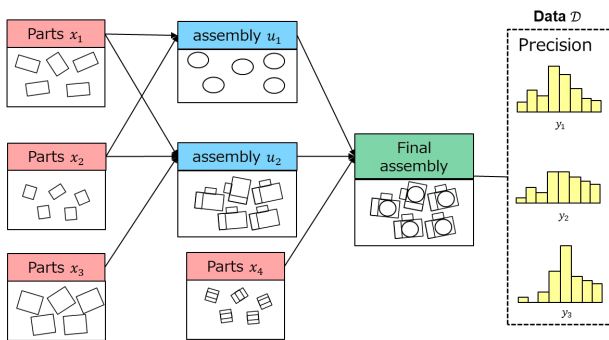


Fig. 1 生産工程の概略図。部品 (parts) x から部組 (assembly) u が組み立てられ、さらにそれらを組み合わせて総組 (final assembly) が完成する。総組の性能が検査され N 個分のデータ y が得られる。

部品パラメータ x は、部品を特徴づける量を表す。例えば、穴径などの部品寸法や設計値からの位置ずれであり、設計仕様に基づいて予め定義域が決められている。これらが組み合わされ、部組といわれる中間的なモジュールが構成される。この特徴量を u と表す。さらにこれらが組み合わされて総組が完成する。設計指針に基づき、総組が検査され性能 y が得られる。

本研究では、総組の生産数 N 台を 1 ロットとし、ロット単位でデータが管理されていることを想定し、取得されるデータ D は 1 ロットあたり $\{y^{(i)}\}_{i=1}^N$ と書く。不良要因検知の目標は、基準となる正常ロットから取得された製品データセット D_A と、不良が生じたロットから取得された製品データセット D_B が与えられた場合に、不良要因となる部品パラメータを推定することである。しかし、特殊な場合を除いて、工程の上流のデータは得られないため、完全なデータ駆動型の変化要因検知は不可能である。そのため本研究では、生産工程シミュレーターと密度比推定を用いた不良要因検知手法を提案する。提案手法は、次の 2 つのパートからなる。

(i). 正常データセット D_A から、生産工程シミュレーターを用いて部品パラメータ x の分布を推定する。

(ii). 上記 (i) で推定された部品パラメータセットの分布に基づいて、検知対象データ D_B から、各部品パラメータの不良要因スコアを計算する。

特に上記 (i). は、生産工程シミュレーターを f として

$$y = f_0(x) + \epsilon \equiv f(x) \quad (1)$$

において、製品性能 y から部品パラメータ x を求める問題に対応する。ここで x と y はそれぞれ D_x , D_y 次元のベクトルとする。また ϵ は工程中に含まれるノイズであり、確率的因子として生産工程シミュレーター f に含まれている。また f_0 はベースラインとなる決定的モデルであり、式 (1) は単体の製品の性能と部品パラメータの関係を表していることに注意する。仮に、工程に含まれるノイズが適切にモデリングされており、取得されるデータが単体のものであれば、式 (1) に何らかの逆問題解法（正則化やベイズ推定）を用いることで、部品パラメータを求めることは可能かもしれない (Fig. 2 (a)). しかし本研究では量産データを扱うため、単体の製品性能が出力であるような式

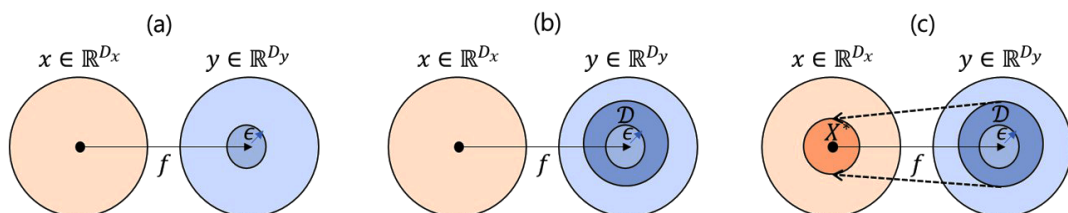


Fig. 2 (a). 生産工程シミュレーターの概念図。 (b). 製品集合の性能データ $D = \{y_i\}_{i=1}^N$ とシミュレーターの関係。 (c). 問題 (i) はデータ集合 D を表現する部品パラメータ集合 X^* を見つけることである。

(1) の逆問題として定式化することは難しい (Fig. 2 (b)). 本研究では, 製品集合の性能から, これを上手く表現するような部品パラメーターの集合 X^* を推定するアルゴリズムを提案する. 次に, 正常データセットと不良品が含まれるデータセット D_B に対して, 上記アルゴリズムを適用し, 差分を調べることで不良の要因候補を見つけることができるはずである.

さらに (i) の過程でデータセット D から推定された X^* を用いることで $y=f(x)$ なる統計モデルを推定することが可能である. これは, シミュレーターと同等の入出力関係を持つエミュレーターと見做すことができる.

3 提案手法

本節では, 提案法の詳細を述べる.

3.1. 近似ベイズ計算を用いた部品パラメーターの推定:

Approximate Bayesian Computation (ABC) inspired inverse

部品パラメーター x の事後分布は, ベイズの定理によって次のように書ける.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \quad (2)$$

ここで $p(y|x)$ は x が与えられた下での y の尤度であり, 式 (1) におけるシミュレーターに対応する $p(x)$ は事前確率に対応し, 部品パラメーターが予め持っている情報を記述することができる. 近似ベイズ計算は, 尤度が複雑な関数である場合でも, データ $D = \{y_i\}_{i=1}^N$ が与えられた下で, 次のようなプロセスで近似的に事後確率を近似的に推定することができる方法論である [6], [7].

1. 事前分布 $\pi(x)$ から N 個のパラメーター $x^{(j)}, j = 1, \dots, N$ をサンプリングする.
2. シミュレーションを行う: $y^{(j)} = f(x^{(j)}), j = 1, \dots, N$.
3. データとシミュレーションの出力の距離

$$\rho(S(D), S(y^{(i)})) \leq \epsilon \quad (3)$$

を計算し, ϵ が小さい $x^{(i)}$ を受理する. ここで, S は十分統計量, ρ は距離尺度である.

上記の通り, 通常の近似ベイズ計算では, シミュレーターの出力がデータの分布全てを表現することを仮定しているが, 本研究で使用するシミュレーターは前節で述べたように製品単体の性能を出力するため, 製品集合全体の分布を表現できない. そこで, 事前分布 $\pi(x)$ の範囲を広く取っておき, N_{sim} 回サンプリングを行い, それぞれの部品パラメーター $x^{(j)}$ に対してシミュレーションを行うことを考える. すなわち, 量産データを表現するために広範囲からサンプリングしたパラメーター x に対してシミュレートして, データ y を包含するような分布を作っておく. シミュ

レーションによって得られる分布は, データを含む大きな分布になるが, ここから密度比

$$r(y) = \frac{p^{\text{data}}(y)}{p^{\text{sim}}(y)} \quad (4)$$

に比例する受理確率に従って x を受理することにする. 設計指針に基づき, 総組が検査され性能 y が得られる.

ここで, $D = \{y_i\}_{i=1}^N \sim p^{\text{data}}(y)$ と $D_{\text{sim}} = \{y^{(j)}\} \sim p^{\text{sim}}(y)$ はそれぞれデータとシミュレーター出力の密度である. このような受理基準は, 次のような性質がある.

- $r(y^{(j)})$ が小さければ, データ密度が小さく, このような場合は $y^{(j)}$ を出力する $x^{(j)}$ は棄却される.
- $r(y^{(j)})$ が大きければ, データ密度が大きく, このような場合は $y^{(j)}$ を出力する $x^{(j)}$ を受理する.

このアルゴリズムを ABC inspired inverse と呼び, 詳細を Algorithm 1 にまとめる. また, 出力されるサンプルに対して, 次の定理が導ける.

Lemma. Algorithm 1 において受理されたサンプル $Y^* = \{y^{*(j)}\}_{j=1}^{n'}$ の分布は $p^{\text{data}}(y)$ 一致する.

Theorem. Algorithm 1 において受理されたサンプル $X^* = \{x^{*(j)}\}_{j=1}^{n'}$ の分布は事後分布 $p(x|D)$ と一致する.

証明は Appendix に掲げる. 実際には, 密度比 r をデータ集合 D とシミュレーターの出力集合 D_{sim} から推定する必要があり, この意味で推定誤差が加わる. 密度比は Kullback-Leibler importance estimation procedure (KLIEP) と呼ばれる手法によって p^{sim} と p^{data} を各々推定することなく, 直接推定することができる [8].

Algorithm 1 ABC inspired inverse

Input: $D = \{y_i\}_{i=1}^N, f(\cdot), \pi(\cdot),$ and N_{sim}

Output: $X^* = \{x^{*(j)}\}_{j=1}^{n'}$

- 1: **for all** $j = 1, \dots, N_{\text{sim}}$ **do**
 - 2: Draw a sample from a prior: $x^{(j)} \sim \pi(x)$.
 - 3: Perform simulation: $y^{(j)} = f(x^{(j)})$.
 - 4: **end for**
 - 5: Compute the density ratio by applying KLIEP: $r(y) = p^{\text{data}}(y)/p^{\text{sim}}(y)$.
 - 6: Accept each sample $x^{(j)}$ with the probability proportional to the density ratio.
-

3.2. 不良要因スコアの計算

ここでは, 正常データ D_A と, 別ロットから取得された検知対象データ D_B が取得された場合を考える. D_B は性能 y に不良があることを想定しているため, これを実現する部品パラメーターセットの分布は, D_A に対して Algorithm 1 を適用して得られるサンプル集合 X_A^* の分布とは異なっていると考えられる. ここで, 不良ロットのデータ D_B は不良要因となる部品パラメーターの影響を強く受けて変化しているとする. 逆に, 不良要因ではないパラメーターは, どのような値をとっても D_B を再現できないと考えられる. このような考えに基づいて, 正常時の部品パラ

メーターの分布を用いて不良要因スコアを導出する。

まず正常データ \mathcal{D}_B に対して Algorithm 1 を適用し、サンプル \mathbf{X}_A^* を得る。次に、不良要因が独立に \mathcal{D}_B の性能 \mathbf{y} に影響を与えらるとして、 x_d を事前分布からサンプリングし、それ以外のパラメーター $\mathbf{x}_{\setminus d}$ を \mathbf{X}_A^* からサンプリングする。このようにサンプリングされた部品パラメーター集合は、特定の変数 x_d 以外は正常データを出力するような集合であることに注意する。このようにサンプリングされた部品パラメーター集合と検知対象データ \mathcal{D}_B に対し、Algorithm 1 を適用し、受理数 a_d を計算する。 a_d が大きければ不良データを再現しており、 a_d が小さければ x_d をどのようにサンプリングしても不良データを再現できない、と考えることができる。従って、受理数 a_d は、 x_d が不良データをどの程度再現するかの指標であり、この値が大きければ x_d が不良要因である可能性が高いと考えられる。ただし、各部品パラメーターの最終的な不良要因スコアを算出する際には、次のことに注意する。

- ベースライン：設計仕様や調整プロセスが異なるため、総組の性能 \mathbf{y} への感度は部品パラメーターに依存してまちまちである。
- 乱数系列依存性：確率的なアルゴリズムのため、乱数系列依存性がある。

上記を踏まえて、部品パラメーター x_d の最終的な不良要因スコア $\{c_d\}_{d=1}^{D_x}$ を次のように定義する。

$$c'_d = \frac{1}{M\bar{a}_d^0} \sum_{m=1}^M a_d^m, (1 \leq d \leq D_x) \quad (5)$$

$$c_d = \frac{c'_d}{\max(\mathbf{c}')} \quad (6)$$

ここで、 a_d^m は m 番目の乱数シードにおいて、 \mathcal{D}_B における x_d の受理数である、また、 $\bar{a}_d^0 \equiv 1/M \sum_{m=1}^M a_d^{m,0}$ で定義され、 \mathcal{D}_A における受理数のシード平均を表す。また、 $\max(\mathbf{c}')$ は $\mathbf{c}' = [c'_1, \dots, c'_{D_x}]$ の最大値を表し、不良要因スコア c_d は 0 から 1 の範囲を取る。不良要因スコア計算のためのアルゴリズムを Algorithm 2 にまとめる。

3.3. エミュレーターの推定

本節では Algorithm 1 によって推定された部品パラメーターのサンプル \mathbf{X}^* とデータ \mathcal{D} を用いて、エミュレーター (疑似モデル) を推定する。生産工程中では特定の部品パラメーターの値が調整されることがあるため、最終的な性能に対して感度がない部品パラメーターが含まれる。

L1正則化付き線形モデルを用いることで、性能を出力する統計モデルを構成することができると共に、最終的に感度がある部品パラメーターを自動的に抽出することができる。

総組の性能 \mathbf{y} と部品パラメーター \mathbf{x} の関係が

$$\mathbf{y}_k = f_k(\mathbf{x}) = \mathbf{w}_k^T \mathbf{x}, 1 \leq k \leq D_y, \quad (7)$$

と表すことができ、係数 $\mathbf{w}_k \in \mathbb{R}^{D_x}$ はスパースなベクトル

Algorithm 2 Defect factor detection via the ABC inspired inverse

Input: $\mathcal{D}_A, \mathcal{D}_B, \mathbf{f}(\cdot), \pi(\cdot), N_{\text{sim}},$ and M

Output: $\mathbf{c} = \{c_d\}_{d=1}^{D_x}$

- 1: Apply the ABC inspired inverse to \mathcal{D}_A , and obtain $\mathbf{X}_A^* = \left\{ \mathbf{x}^{*(j)} \right\}_{j=1}^{n'}$.
- 2: **for all** $m = 1, \dots, M$ **do**
- 3: **for all** $\xi = \mathcal{D}_A$ and \mathcal{D}_B **do**
- 4: **for all** $d = 1, \dots, D_x$ **do**
- 5: **for all** $j = 1, \dots, n'$ **do**
- 6: Generate samples: $x_d^{(j)} \sim \pi(x_d), \mathbf{x}_{\setminus d}^{(j)} = \mathbf{x}^{*(j)}$.
- 7: Perform simulator: $\mathbf{y}^{(j)} = \mathbf{f}(\mathbf{x}^{(j)})$.
- 8: **end for**
- 9: Compute the density ratio by using KLIEP: $r(\mathbf{y}) = p^{\text{data}}(\mathbf{y})/p^{\text{sim}}(\mathbf{y})$.
- 10: Accept each sample $\mathbf{x}^{(j)}$ with the probability proportional to the density ratio.
- 11: Assign the total number of accepted samples $a_d^{m,\xi}$.
- 12: **end for**
- 13: **end for**
- 14: **end for**
- 15: Calculate the defect score (6).

であると仮定する。すなわち、総組の精度と部品パラメーターの関係を完全な線形モデルで近似し、なおかつ影響を及ぼす部品パラメーターは限られているとする。データ \mathcal{D}_A から Algorithm 1 によって推定された部品データの集合 $\mathbf{X}_0 = \{\mathbf{x}_0^{(i)}\}_{i=1}^N$ から $\{\mathbf{w}_k\}_{k=1}^{D_y}$ を次のように推定する。

$$\mathbf{w}_k^* = \arg \min_{\mathbf{w}_k} \left[\sum_i \left(y_k^{(i)} - \mathbf{w}_k^T \mathbf{x}_0^{(i)} \right)^2 + \lambda \|\mathbf{w}_k\|_1 \right], \quad 1 \leq k \leq D_y, \quad (8)$$

ここで $\|\cdot\|_1$ は L1 ノルムを表し、 λ は L1 正則化の強さを決めるパラメーターである。この最適化は Lasso と呼ばれる Least Angle Regression (LARS) や Alternating Direction Method of Multipliers (ADMM) を用いて実行できる [9], [10].

4 シミュレーターの構成

本研究では、カメラ内部のモジュールの生産を扱う。このモジュールの生産工程は、Fig. 3 のような工程モデルによって表される。16 の部品 (x_1, \dots, x_{16}) から部組が構成され、そのうちの 3 つの部品 (x_1, x_5, x_{12}) に関して調整を行い、さらに 4 つの部品 (x_{17}, \dots, x_{20}) を組付け、総組状態で最終的に 4 次元の精度指標 (y_1, \dots, y_4) が計測される。すなわち $D_x = 20, D_y = 4$ である。なお、調整工程では u_i, u'_i, u''_i, u'''_i などの中間出力が一定の規格内に収まるよう部品 (x_1, x_5, x_{12}) の位置を $\Delta x_1, \Delta x_5, \Delta x_{12}$ だけずらす。量産工程では 500 の総組 1 ロットとして生産されるとする。

部品パラメーター x_d の事前分布は、一様分布 $U(\lambda_d^l, \lambda_d^u)$ で与えられることとし、分布のパラメーターは設計仕様か

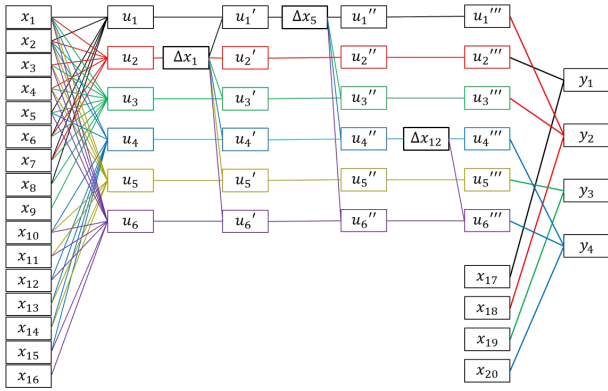


Fig. 3 カメラ内モジュールの工程モデル。20の部品パラメータから総組が構成され、4つの精度指標で検査が行われる。本工程では、 x_1 、 x_5 と x_{12} に対して調整が行われる。

ら次のように設定する。

$$\lambda_d^1 = \begin{cases} -0.1, 1 \leq d \leq 16 \\ -0.2, 17 \leq d \leq 20 \end{cases}, \lambda_d^2 = \begin{cases} 0.1, 1 \leq d \leq 16 \\ 0.2, 17 \leq d \leq 20 \end{cases} \quad (9)$$

5 結果

5.1. 不良要因検知

5.1.1. 数値シミュレーション

本節では、不良要因検知のシミュレーションを行い、提案手法の有効性と限界を示す。これは次のような手続きで行う。

1. 部品パラメータ x_d を正規分布 $\mathcal{N}(\mu_d, \sigma_d^2)$ に従って N サンプル生成し、 $\mathbf{y}^{(i)} = \mathbf{f}(\mathbf{x}^{(i)})$, $i=1, \dots, N$ を計算する。このセット $D_A = \{\mathbf{y}^{(i)}\}_{i=1}^N$ を正常データとする。ここで、

製品仕様から

$$\mu_d = 0, 1 \leq d \leq 20, \sigma_d^2 = \begin{cases} 0.01, 1 \leq d \leq 16, \\ 0.02, 17 \leq d \leq 20 \end{cases} \quad (10)$$

と決めた。

2. 不良要因は正常の分布からずれていると仮定し、正規分布 $\mathcal{N}(\mu_f + \delta\mu_f, (\sigma_f + \delta\sigma_f)^2)$ から N サンプル生成する。その他の部品パラメータ x_d は正常時の分布 $\mathcal{N}(\mu_d, \sigma_d^2)$ から N サンプル生成する。このサンプルセットを用いて $\mathbf{y} = \mathbf{f}(\mathbf{x})$ から $D_B = \{\mathbf{y}^{(i)}\}_{i=1}^N$ を検知対象データとする。

3. D_A と D_B を用いて Algorithm 1 と Algorithm 2 を適用し、最終的な不良要因スコア $\{c_d\}_{d=1}^{D_x}$ を得る。

まず、下流部品パラメータ x_{19} を不良要因だと想定し、生成の平均と分散を $\delta\mu_{19} = 0.1$, $\delta\sigma_{19} = 0.1$ だけ変化させた場合を Fig. 4 (a) 左に示す x_{19} が変化したことによって y_3 が変化している。Fig. 4 (a) 右から、真の不良要因パラメータ x_{19} のスコア c_{19} が高い値になっていることが分かる。部品パラメータ x_{13} , x_{14} と x_2 も高いスコアを示しているが、これは y_3 と関連しているパラメータであることが Fig. 3 からわかるため、合理的な結果であると考えられる。Fig. 4 (b) は、 x_1 を変化させた場合の結果である。この場合、性能パラメータ \mathbf{y} と不良要因スコアに顕著な差異が見られないが、これは x_1 が調整工程を経ているため \mathbf{y} に対して感度がほとんどないことが原因であると考えられる。このように、結果系に対して感度が低い変数は不良要因として特定することは不可能である。一方、Fig. 4 (c) は x_7 を真の不良要因だと想定した場合である。上流のパラメータであっても c_7 が顕著に高い値を示し、不良要因として特定で

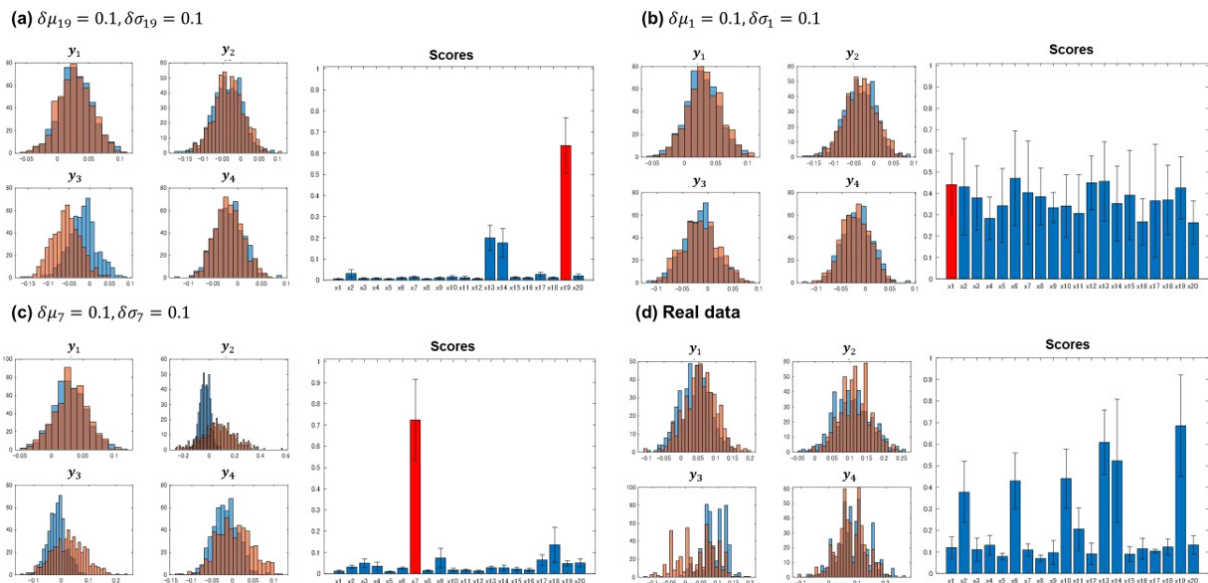


Fig. 4 (a)-(c). シミュレーションの結果。左列のヒストグラムは D_A の分布 (青) と D_B の分布 (赤) を表す。右列の棒グラフは不良要因スコア c を表し、エラーバーは標準偏差を表す。右列は不良要因スコアを表し、真の不良要因を赤で表している。(d). 左列のヒストグラムは実際に取得された正常ロットのデータ D_A の分布 (青) と不良品を含むロットのデータ D_B の分布 (赤) を表す。

きていることが分かる。

5.1.2. 実データの検証

本節では、実際の生産現場で取得されたデータの解析結果を述べる。取得された約140ロットの生産データから、正常に生産されているロットと不良が確認されたロットを1つずつ選定した。Fig. 4 (d) にデータと不良要因スコアを示す。このときは、特に y_3 が顕著に変化しており、不良要因スコア c_{19} が顕著に高い値を取っていることが分かる。実際の現場では x_{19} の部品に不具合があったことが確認されていることから、これは得られた結果と整合する。また、その他の高い不良要因スコアを持つ変数についても、Fig. 3 から y_3 に紐づく変数が検出されていることから、製品設計の観点から自然な結果であると考えられる。

5.2. エミュレーターの推定

Fig. 5 (a) は、設計の感度解析から得られる y と x の設計知識による感度解析の結果である。0 は感度が低いことを表し、1 は感度が高いことを表す。Fig. 5 (b) は、 y の各々の成分に対して推定された w^* である。横軸は正則化パラメーター λ を表し Fig. 中の垂線はクロスバリデーションの結果を表す。 w^* は20次元のベクトルであり、Fig. 5 (a) のドメイン知識に基づいて、感度があるものは実線、ないものは破線で表現している。この図から、感度がない部品パラ

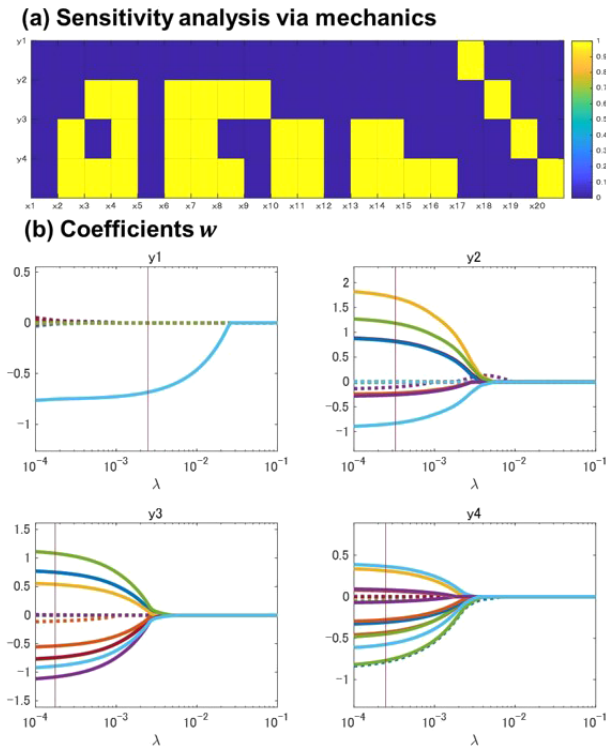


Fig. 5 (a). 設計の感度解析からドメイン知識として得られる結果。精度 y_k に感度が高ければ1、低ければ0を表している。(b) : Lassoによって得られた正則化パラメーター λ に対する係数 w の結果。垂線はクロスバリデーションで得られた λ の値である。

メーターに対応する成分は0になり、感度があるパラメーターは有限の値を取る傾向があることが分かる。したがって、完全な線形モデルで近似された式 (7) で表されるエミュレーターは、性能と部品パラメーター間の感度を表すモデルとして有用であると考えられる。また w^* が0になりやすい部品パラメーターは、性能への感度が小さいため、不良要因スコアとして抽出しづらいといえる。

6 まとめ

本論稿では、限られたデータしか取得できないような生産工程において、シミュレーションを用いた逆問題解法 (ABC inspired inverse) と不良要因検知手法を提案した。シミュレーションを通じて提案手法の適用可能性と限界について論じ、検知可能な変数の特性を明らかにした。さらに、実際の生産工程で得られたデータを用いて提案手法を適用し、現場での検証結果と整合する結果を得ることができた。また、複雑な非線形モデルを代替するエミュレーターの構成方法を提案した。提案手法はスパースモデリングを活用することで、感度のある部品パラメーターを自動で選択することも可能である。本研究の成果は、生産工程の効率化・自律化の観点で重要な役割を果たすと考えられる。

Appendix A. Lemma の証明

次の表記を用いると

$$I = \begin{cases} 1 & \text{if } y \text{ is accepted} \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A.1})$$

受理された y の分布は

$$\begin{aligned} p(y^*) &= p(y|I=1) = \frac{p^{\text{sim}}(y)\Pr(I=1|y)}{\Pr(I=1)} \\ &= \frac{c p^{\text{sim}}(y) \frac{p^{\text{data}}(y)}{p^{\text{sim}}(y)}}{\Pr(I=1)} = p^{\text{data}}(y), \end{aligned} \quad (\text{A.2})$$

と計算できる。ここで $\Pr(I=1|y)$ はサンプルの受理確率を表し、式 (4) よりデータの出力とシミュレーターの出力の密度比に比例する。比例定数は $c \equiv \Pr(I=1)$ としている。

□

Appendix B. Theorem の証明

$$\begin{aligned} p(x|I=1) &= \int p(x|y) p(y|I=1) dy \\ &= \pi(x) \int \frac{p(y|x) p^{\text{data}}(y)}{p^{\text{sim}}(y)} dy \end{aligned} \quad (\text{A.3})$$

ここで Lemma を用いた。一方、ベイズの定理より x の事後確率は下記のように表すことができる。

$$p(\mathbf{x}|\mathcal{D}) = \frac{\pi(\mathbf{x})}{p(\mathcal{D})} \int p(\mathbf{y}|\mathbf{x}) p(\mathcal{D}|\mathbf{y}) d\mathbf{y} \quad (\text{A.4})$$

よって、式 (A.3) と式 (A.4)、 $p(\mathcal{D}|\mathbf{y}) \propto \frac{p^{\text{data}}(\mathbf{y})}{p^{\text{sim}}(\mathbf{y})}$ により $p(\mathbf{x}|I=1) = p(\mathbf{x}|\mathcal{D})$ となり、受理された \mathbf{x} の分布と事後分布が一致することが示された。□

引用文献

- [1] B. B. a. H. K. J. Lee, "A Cyber Physical Systems architecture for Industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18-23, 2015.
- [2] 岡崎隼也, "半導体製造工程における Virtual Metrology 技術," *応用数理*, 第29巻, 第1, pp. 31-34, 2019.
- [3] J. M. a. J. Iskandar, "Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing," *Processes*, vol. 5, no. 39, 2017.
- [4] 佐々木豊春, 新海雅彦, 東山孝一郎, 田中文基, 岸浪健史, "鏡筒光学製品における統計的公差設計システム," *精密工学会誌*, 第64巻, 第7, pp. 1090-1095, 1998.
- [5] I. R. a. G. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*, CRC Press, 2015.
- [6] Y. F. a. M. A. B. S. A. Sisson., *Handbook of approximate Bayesian computation*, Chapman & Hall/CRC., 2016.
- [7] D. J. B. R. C. G. a. P. D. S. Tavare, "Missing data: Our view of the state of the art," *Psychological Methods*, vol. 7, no. 2, pp. 147-177, 1997.
- [8] T. S. a. T. K. M. Sugiyama, *Density Ratio Estimation in Machine Learning*, Cambridge Press, 2012.
- [9] R. Tibshirani, "Regression Shrinkage and Selection via Lasso," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267-288, 1996.
- [10] S. Boyd, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2010.
- [11] 大坪洋介, 大谷直也, 近末恵美, 杉山将, "産工程における不良要因検知," *第34回人工知能学会全国大会論文集*, Vols. 214-GS-2-04, pp. 1-4, 2020.

大坪洋介 Yosuke OTSUBO
研究開発本部 数理技術研究所
Mathematical Science Research Laboratory
Research & Development Division

大谷直也 Naoya OTANI
研究開発本部 数理技術研究所
Mathematical Science Research Laboratory
Research & Development Division

近末恵美 Megumi CHIKASUE
研究開発本部 数理技術研究所
Mathematical Science Research Laboratory
Research & Development Division

杉山 将 Masashi SUGIYAMA
理化学研究所
RIKEN
東京大学
The University of Tokyo



大坪洋介
Yosuke OTSUBO



大谷直也
Naoya OTANI



近末恵美
Megumi CHIKASUE



杉山 将
Masashi SUGIYAMA