入門講座

インフォマティクス入門－3

材料工学における機械学習による順解析と逆解析

Direct Analysis and Inverse Analysis by Machine Learning in Material Science and Engineering

名古屋大学
大学院工学研究科
材料デザイン工学専攻　教授
足立吉隆
Yoshitaka Adachi

名古屋大学
大学院工学研究科
材料デザイン工学専攻　講師
小川登志男
Toshio Ogawa

名古屋大学
大学院工学研究科
材料デザイン工学専攻　助教
Zhi-Lei Wang

1. 緒言

材料組織と特性を関連付けるとき、イメージベースのモデリング、因果関係に基づく手法、相関関係を最大限使う方法がある（Fig.1）。イメージベースモデリングは複雑な組織であっても特性の予測ができるが一般的に計算負荷が大きい。因果関係に基づく手法は、比較的シンプルな組織に向いた手法であり結晶粒径などの材料工学的に重要な特徴量と特性との関係を一般化された式を使って関連付ける。単独の組織特徴量の影響を評価する場合にはこの因果関係による手法は理論的背景を伴っているので好ましい手法といえるが、現象が重畳する場合の加算則が必ずしも明らかではなく複雑系での適用には限界がある。複雑系では相関関係を使う手法が有用であり、そこでは主にデータサイエンスを使う。材料組織の特徴量としては、前報[1,2]で述べた材料組織形態の特徴量に加えて、結晶学的特徴や化学組成あるいは各相の硬さなどの物性値を加え、そこに特性をデータとして加えてデータベースを作成すると、機械学習に入力する準備が整う。入力データは必ずしも実験データだけではなく、セルオートマトン、フェーズフィールド法などのシミュレーションや作図により描いたバーチャル組織、マイクロメカニクスや有限要素法あるいは理論により求めた特性データも対象である。すなわち機械学習は実験、理論、シミュレーションの結果を統合する便利な手法である。

抽出したすべての特徴量が特性に影響するとは限らず、重要な特徴量を見極める必要がある。入力変数と出力変数の関

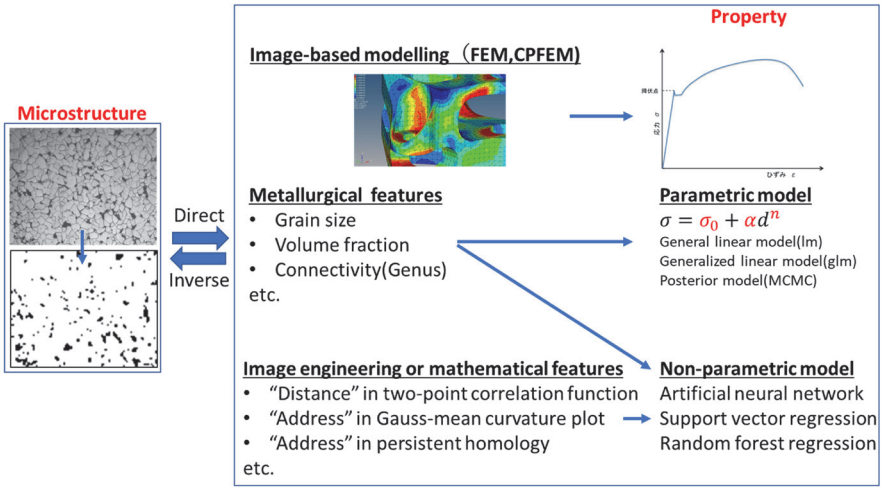

Fig.1　A route to predict a property.

係を最適にモデル化する際にはこの過程が重要となり、スパース学習[3]の重要な部分を占めている。機械学習ではテストデータに対しては一見良いモデルができているように思えても、新規データに対してはそのモデルによる予測結果が満足いかないといった過学習という問題がしばしば生じる。この過学習を抑制するうえで、入力変数の削減は重要である。重要な特徴量をデータサイエンスでは記述子と呼ぶが、記述子を選択する手法（Fig.2）として、赤池情報量規準（AIC）[4]，ベイズ的情報量規準（BIC）[5]，lasso[6]，感度解析[7]などの手法がある。

入力変数を削減する際には上述したように記述子を選択するという手法以外に、入力変数の次元を圧縮する手法がある。この手法としては、主成分分析（PCA）[8]やオートエンコーダー[9]などがあり、非線形の主成分分析を行うカーネル主成分分析[10]といった手法もある。

記述子を選択する、あるいは次元圧縮を行って入力変数を過不足なく削減した後に、記述子と特性（場合によってはトレードオフ特性バランス）の関係を究極的な近似式で表現することを順解析（Fig.2）と呼ぶ。究極的な近似式で表現するということは、因果関係（物理的な意味を持った関係）に基づいた物理モデルを作るというわけではなく、相関関係を最大限使った回帰モデルを作るということを意味する。物理的背景は考慮していないが、そのモデルの精度は特にデータにノイズがある場合に物理モデルよりも優れる場合がある。順解析モデルとしては、汎用されている重回帰線形モデル（RA）に加えて、非線形な関係がモデル化できるガウス過程回帰（GPR）[11]、非線形伝達関数と中間層という概念を取り入れて入出力関係の表現力を高めたニューラルネットワーク回帰（ANN）[12]、プロットのマージン（重み係数の逆数）を最大化するところに回帰直線を引き更にカーネル法により非線形関係まで取り扱えるように工夫されたサポートベクターマシン回帰（SVR）[13]、複数の決定木（正確には回帰木）の結果の平均をとるランダムフォレスト回帰（RF）[14]などが代表的な機械学習法である。いずれの機械学習法にもパラメータがあり、それらをハイパーパラメータという。過学習を抑えかつ精度の良いモデルを作るためにはハイパーパラメータの最適化が必要である（詳細は後述する）。

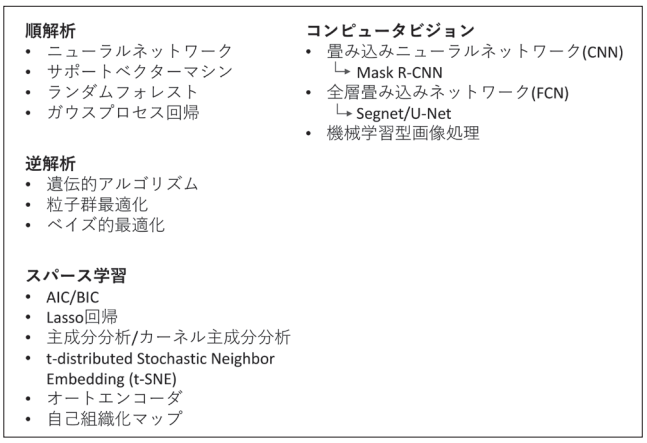

Fig.2　A variety of machine learning methods.

各種順解析モデルは基本的には四則演算で究極的な近似式を作っているだけであるので、計算が軽いことは次に述べる全数探索的な計算を行う逆解析（Fig.2）では重要なことである。

良い順解析モデルができたならば、その順解析モデルを使って入力変数の組み合わせを変えながら出力変数を算出することを全数探索的に行えば、順解析モデル作成時に用いた出力変数のデータの値を超える出力変数が得られる可能性がある。ここで順解析モデルがモデル作成時に用いた入力変数の範囲のみその精度が保証されていることを考慮すると、逆解析の精度を保証できるのはその入力変数の範囲のみであることに注意する必要がある。一方、出力変数は元の出力変数の範囲を超えても構わない。ただし、実際に無限に入力変数の組み合わせを変えて逆解析することは非効率的であり時間がかかりすぎる。そこで出力結果がより大きく（あるいはより小さく）なる方向に入力変数の組み合わせを効率よく変化させる工夫が必要である。その工夫がなされた逆解析手法として、本稿では遺伝的アルゴリズム（GA）[15]、粒子群最適化（PSO）[16]、ベイズ的最適化（BO）[17]を取り上げる。GA, PSOは比較的データ数が多い時に最適値を探索することに向いており、一方BOはデータ数が少ない時にできるだけ少ない実験やモデリングで最適値を探索する際に用いられる。

なお、本稿では主に特性を最大化する組織を提示することを逆解析と呼ぶが、これ以外にも実験データからモデル式のパラメータを最適化することも逆解析であり、手法は共通しているが詳細は別報[18]を参考にしていただきたい。逆解析では、1つの目的変数あるいは2つの目的変数の積や和を最大化する入力変数を探索するが、目的変数が3つ以上の場合それらを同時に最大化する入力変数を探索するには工夫が必要である。これは、特性あるいは二つのトレードオフの関係にある特性のバランスを最大化する最適な組織特徴量（の組み合わせ）を探索することは比較的容易であるが、その最適な組織特徴量を実現するプロセス・組成の探索は容易ではないことを意味する。この問題を解決するためにWang[27]は、組織特徴量とプロセス・組成を同一階層で並列化し、目的変数である特性を最大化する組織特徴量、プロセス、組成を同時に逆解析する手法を行っている。この考え方は、特性を決めているのは第一義的にプロセス・組成であり、特性のばらつきの要因が組織というノイズであるという立場をとっている。ただしこの手法では、組織とプロセス・化学組成が矛盾しないという保証はなく、特性➡組織➡プロセス・組成を階

層的に最適化するには多くの課題が残されている。

各機械学習のモジュールはプログラミング言語PythonやRで公開されており、今日では比較的容易に機械学習の計算ができるが、その基本的なことを理解することは重要と思われるので、本稿では必ずしも機械学習を専門としない材料工学関係者が機械学習の理解を深めていただくことを主眼に執筆する。

2 スパース学習

2.1 重要な入力変数の選択

最適な順解析モデルを作るにあたり、出力変数に影響を与える程度が大きい入力変数を選択することが過学習を抑制するうえで必要である。専門家がその重要性を認識しているのでその知識に基づいて選択する場合もあるが、ここではコンピュータが重要な入力変数を選択する手法について述べる。なお、入力変数の重要性を評価する一つの手法が感度解析であるが、これについては3.1.2で述べる。

2.1.1 赤池情報量規準（AIC）、ベイズ情報量規準（BIC）

AIC[4]は、汎化誤差が最も小さくなる予測能力を持つモデルが最良のモデルと考えるのに対し、BICは真のモデルである確率が最も大きいモデルを良いと考える。AICでは、入力変数と出力変数の関係を重回帰分析した際に、尤度（L）が最大になることを基本としてそこに説明変数の数（k）の要素も考慮して最適モデルを選択する。AICの値が最も小さくなるモデルが最適であると考える。Fig.3（a）の二相組織鋼の例では、出力変数の強度（s）に及ぼす入力変数の1. フェライト相の硬さ（HvF）、2. マルテンサイト相の硬さ（HvM）、3. マルテンサイト相の体積率'VMM）、4. マルテンサイト相の数密

(a)AIC

```
Start:  AIC=-218.68
datasetAIC...OutputVariablesAIC.1.. ~ HvF + HvM + VMM + NumDensity +
    tunnel + void + strain

            Df Sum of Sq    RSS     AIC
- NumDensity  1    0.2399 20.252 -219.16
- HvF         1    0.2496 20.261 -219.10
<none>                   20.012 -218.68
- tunnel      1    0.5503 20.562 -217.23
- void        1    0.9726 20.984 -214.65
- VMM         1    5.6856 25.697 -188.92
- HvM         1   18.9899 39.002 -135.93
- strain      1   27.7610 47.773 -110.17

Step:  AIC=-219.17
datasetAIC...OutputVariablesAIC.1.. ~ HvF + HvM + VMM + tunnel +
    void + strain

        Df Sum of Sq    RSS     AIC
<none>                20.252 -219.16
- HvF     1    0.6031 20.855 -217.44
- tunnel  1    0.7007 20.952 -216.84
- void    1    1.2843 21.536 -213.36
- VMM     1   14.4098 34.661 -152.92
- HvM     1   19.3014 39.553 -136.15
- strain  1   27.5573 47.809 -112.08
```

(a)BIC

```
Start:  AIC=-195.93
datasetBIC...OutputVariablesBIC.1.. ~ HvF + HvM + VMM + NumDensity +
    tunnel + void + strain

            Df Sum of Sq    RSS      AIC
- NumDensity  1    0.2399 20.252 -199.256
- HvF         1    0.2496 20.261 -199.196
- tunnel      1    0.5503 20.562 -197.325
<none>                   20.012 -195.925
- void        1    0.9726 20.984 -194.742
- VMM         1    5.6856 25.697 -169.011
- HvM         1   18.9899 39.002 -116.025
- strain      1   27.7610 47.773  -90.263

Step:  AIC=-199.26
datasetBIC...OutputVariablesBIC.1.. ~ HvF + HvM + VMM + tunnel +
    void + strain

        Df Sum of Sq    RSS      AIC
- HvF     1    0.6031 20.855 -200.374
- tunnel  1    0.7007 20.952 -199.780
<none>                20.252 -199.256
- void    1    1.2843 21.536 -196.291
- VMM     1   14.4098 34.661 -135.852
- HvM     1   19.3014 39.553 -119.086
- strain  1   27.5573 47.809  -95.011

Step:  AIC=-200.37
datasetBIC...OutputVariablesBIC.1.. ~ HvM + VMM + tunnel + void +
    strain

        Df Sum of Sq    RSS     AIC
- tunnel  1    0.3529 21.208 -203.09
<none>                20.855 -200.37
- void    1    2.1560 23.011 -192.72
- VMM     1   14.0510 34.906 -139.80
- HvM     1   22.0867 42.942 -113.49
- strain  1   28.8015 49.656  -95.04

Step:  AIC=-203.09
datasetBIC...OutputVariablesBIC.1.. ~ HvM + VMM + void + strain

        Df Sum of Sq    RSS      AIC
<none>                21.208 -203.086
- void    1     2.621 23.828 -193.134
- VMM     1    17.868 39.076 -130.315
- strain  1    29.585 50.793  -97.009
- HvM     1    32.013 53.221  -91.080
```

Fig.3　An example of analysis result by AIC and BIC.

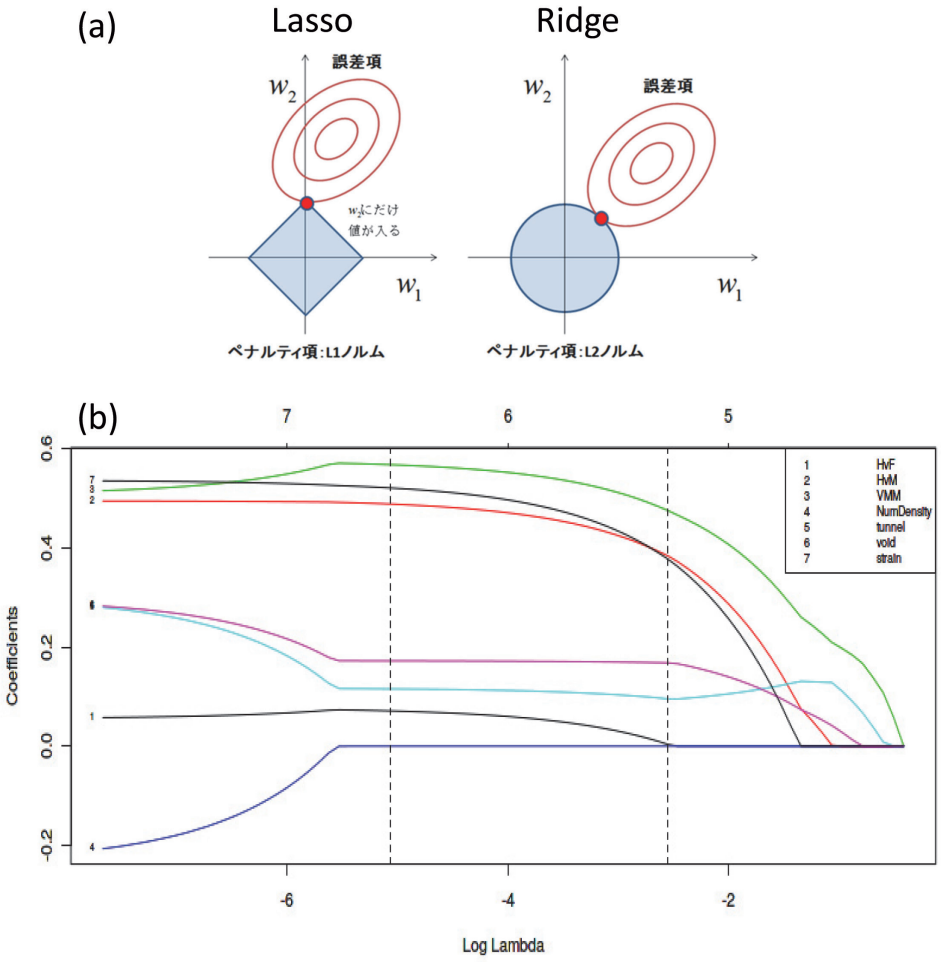

Fig.4　An example of analysis result by lasso regression.

度 (f)、5.マルテンサイト中に貫通している穴の数 (h)、6.同様に空洞の数 (v)、そして7.ひずみ (e) の影響を解析している。データセット数は128個である。

$$AIC = -2lnL + 2k \quad \cdots\cdots (1)$$

fを入力変数から除いた方がAICの値が小さくなるため、二回目の計算ではf以外を入力変数として再計算している。そして、二回目採用した入力変数のいずれを除いてもAICの値が大きくなることから、今回のデータではHvF, HvM, VMM, h, v, eが重要であるという判断に至る。

BIC[5] はAICよりも入力変数を選択する傾向が強い。BICの評価では尤度、説明変数の数に加えて、データ数を考慮している。このBICの値が最も小さくなるモデルが最適と考える。

$$BIC = -2lnL + k * ln(n) \quad \cdots\cdots (2)$$

AICで用いた同じデータに対してBIC評価 (Fig.3 (b)) をすると、最終的にはHvM, VMM, void、eのみが重要であるという判断である。

2.1.2　lasso回帰

lasso回帰[6] では、次式を最小化する重み係数ωを見つける。

$$\frac{1}{2N}\sum_{i=1}^{N}(y_i - \omega_0 - x_i^T\omega)^2 + \lambda P_\alpha(\omega) \quad \cdots\cdots (3)$$

ここで、y_iは実測の出力変数、$\omega_0 + x_i^T\omega$はモデル回帰式、

$$P_\alpha(\omega) = (1-\alpha)\frac{1}{2}\|\omega\|^2 + \alpha\|\omega\| = \sum_{j=1}^{p}[\frac{1}{2}(1-\alpha)\omega_j^2 + \alpha|\omega_j|] \quad \cdots\cdots (4)$$

である。$\alpha = 1$ではlasso回帰、$\alpha = 0$ではリッジ回帰、その他ではElastic Net回帰になる。ここで式 (3) の第一項は誤差項、第二項はペナルティ項である。Fig.4 (a) の例 (ここでは簡単化のため変数は2つとしている) に示すようにlasso回帰ではペナルティ項が一次であるので誤差項と一点で交わりやすく、今回の場合$\omega_1 = 0$となりω_2のみが値を持つ。一方リッジ回帰ではペナルティ項が二次であるので誤差項と至る所で接しω_1, ω_2ともに値を持ち、なかなかゼロにすることができない。このことから、lasso回帰によって、λを変えた時に重み係数ωがいち早くゼロになるものが出てきて、それらの

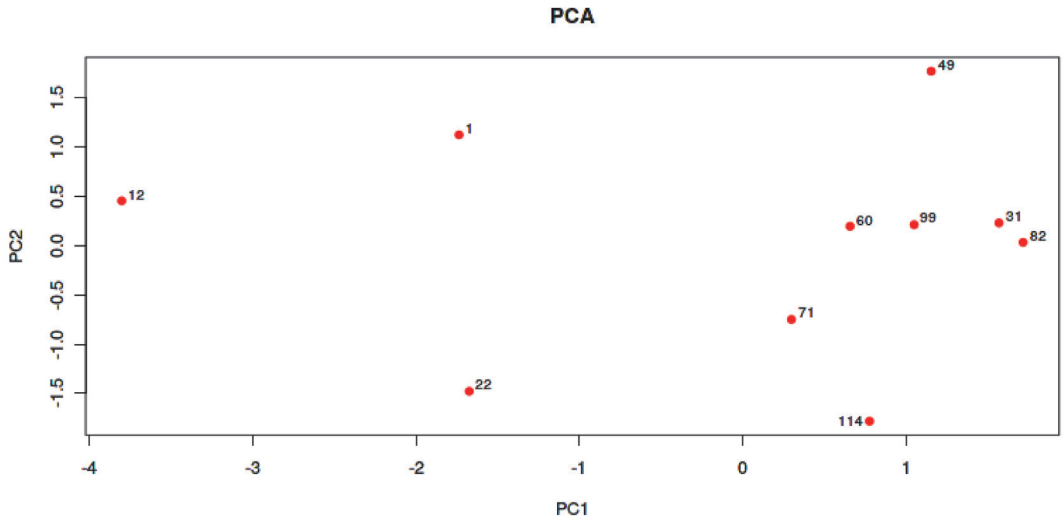

Fig.5　An example of PCA analysis.

重み係数を持つ入力変数は影響力が小さく、記述子からは除かれるべきという判断に至る。先に使った二相組織鋼のデータをここでも使ってlasso回帰した結果をFig.4 (b) に示す。1.HvFや4.NumDensity (f) はいち早くゼロになっており、今回用いたデータ範囲内では重要ではないことが示唆される。この結果は先ほど述べたAIC, BICの結果とも矛盾しない。

2.2　入力変数の次元圧縮

2.2.1　主成分分析 (PCA)

上述した入力変数の選択では例えば7つある入力変数の中から重要と判断された数個の変数のみを記述子として採用し、その他の変数は廃棄した。しかしその廃棄した入力変数が全く影響していないというわけではなく、その影響度を少しでも入力変数として残したい時には、次元圧縮という手法がある。ここではその代表例である主成分分析 (PCA) [8] とオートエンコーダー[9] について述べる。

PCAでは、高次元空間のデータを分散が最大になるように主軸を見つけて低次元空間にプロットする (Fig.5)。二次元空間にプロットする場合、その主軸が第一主成分軸 (PC1) であり、それと垂直成分の軸が第二主成分軸 (PC2) である。この第一主成分軸は固有ベクトルとして求めることができる。

固有値、固有ベクトルとは、行列Aで変換しても、回転せず、単に長さだけが変わるような引き伸ばしのみを行うλとAのことである。今、共分散が次式で与えられるものとすると、共分散行列*1 ($\Sigma=\begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix}$) は以下のようになる。nはデータ数である。

$$S_{xy}=\frac{1}{n}(x-\bar{x})^{T}(y-\bar{y}) \quad \cdots\cdots (5)$$

$$\begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}=\lambda\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad \cdots\cdots (6)$$

ここで$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$は$x=a_1$, $y=a_2$の固有ベクトル、λは固有値である。$a_1^2+a_2^2=1$という制約がある。n次正方行列では固有値、固有ベクトルはn個存在するが、上の例では2×2正方行列なのでそれぞれ2つある。ここで大きい方の固有値の固有ベクトルが第一主成分軸、小さい方に対応するのが第二主成分軸で、二つ軸は直行する。主成分軸に物理的意味はないが、いわば多次元情報が圧縮された新たな特徴量と言える。

通常の主成分分析では妥当な主成分軸を見つけられない場合、元データ群を一度カーネルを使って高次元空間に射影し、そこでより妥当な主成分軸を見つけた後に元の次元の空間に戻す手法をカーネル主成分分析という。どうしても線形射影では妥当な主成分軸を見つけられない時に使える非線形主成分分析である。

主成分分析では似ていないデータは次元圧縮後にできるだけ遠くにプロットする思想に基づいているのに対し、類似しているものは近くにプロットする思想を持つ手法としてt-SNEがあり、次でその要点を述べる*2。なお、前報で述べた自己組織化マップ (SOM) も類似しているデータを近くにプロットする点ではt-SNEと同様である。

＊1　分散共分散行列ともいう。
＊2　その発展版として、極近年umapが報告されている。umapでは、新たな高次元データをすでに次元圧縮した主成分軸のマップにプロットできるという特徴を持つ。

2.2.2 t-distributed stochastic neighbor embedding (t-SNE)

PCAではデータの分散ができるだけ大きくなるような主軸を見つけており、これは似ていないデータはできるだけ遠くにプロットすることになる。この思想に対して、t-SNE[19]では、高次元空間におけるデータ (x_i) 同士の「近さ (類似度)」が、低次元空間におけるデータ (y_i) 同士の「近さ」に反映されるよう学習を行う。

高次元空間で、x_iからみたx_jが存在する確率を、正規分布 (標準偏差：σ) を仮定して、次式で与える。

$$p_{j|i}\frac{\exp\left(-\left\|x_i-x_j\right\|^2/2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\left\|x_i-x_k\right\|^2\right)/2\sigma_i^2},\ p_{i|i}=0 \quad \cdots\cdots (7)$$

ここで$p_{j|i}\neq p_{i|j}$であるが、これを対称化するために、

$$p_{ij}=\frac{p_{j|i}+p_{i|j}}{2N} \quad \cdots\cdots (8)$$

とする。

SNEでは低次元空間でのy_iからみたy_jが存在する確率$q_{j|i}$を高次元空間と同様に正規分布を仮定して設定するが、改良型のt-SNEでは自由度1のt分布を用いる。

$$q_{ij}=\frac{\left(1+\left\|y_i-y_j\right\|^2\right)^{-1}}{\sum_{k,l,k\neq l}\left(1+\left\|y_k-y_l\right\|^2\right)^{-1}},\ q_{i|i}=0 \quad \cdots\cdots (9)$$

正規分布に比べて、t分布はピーク近傍では鋭く、裾野では広がりが大きいので、高次元空間で近くにあるプロットは低次元空間ではより近くに、遠くにあるプロットはより遠くにプロットである特徴がある (Fig.6)。

高次元空間と低次元空間を関連付けるということは、高次元空間でのデータの近さを低次元空間でのデータの近さに反映するということであり、これはp_{ij}とq_{ij}の確率を近づけてやればよいということになる。この二つの確率の距離を測る指標として、次式のカルバック・ライブラリー情報量 (C) が用いられる。このCをできるだけ小さくするようにy_iの座標を更新する。

$$C=\sum_i\sum_j p_{ij}log\frac{p_{ij}}{q_{ij}} \quad \cdots\cdots (10)$$

2.2.3 オートエンコーダー (AE)

主成分分析では数学的にデータを新たな空間に単純射影して主成分軸を見つけているが、機械学習により主成分軸を見つける手法としてオートエンコーダー[9]がある (Fig.7)。オートエンコーダーは第3章で述べるニューラルネットワークを使った手法であるが、通常ニューラルネットワークでは出力変数があるが、オートエンコーダーでは入力変数＝出力変数であり、中間層中のノードを入力変数の数よりも少なくすることにより一度次元を圧縮して、その圧縮した次元のデータをもとのデータに戻すことにより、中間層で得たノードを主成分軸 (AE1, AE2) とする手法である。入力層から中間層へは通常非線形関数を使うので、このオートエンコーダーによる手法も一種の非線形主成分分析と言える。

2.2.4 類似性の指標としての距離

多次元空間のプロットを、低次元空間にプロットすると、類似した特徴を持つプロットは近くに集まる。この時のi-jプロット間の距離については、以下のユークリッド距離が通常使われる。

$$d_{i,j}=\sqrt{\sum_{k=1}^{m}(x_k^{(i)}-x_k^{(j)})^2} \quad \cdots\cdots (11)$$

Fig.8中のa点、c点はそれぞれユークリッド距離で評価するとB, Aグループに属することは明らかであるが、b点はユー

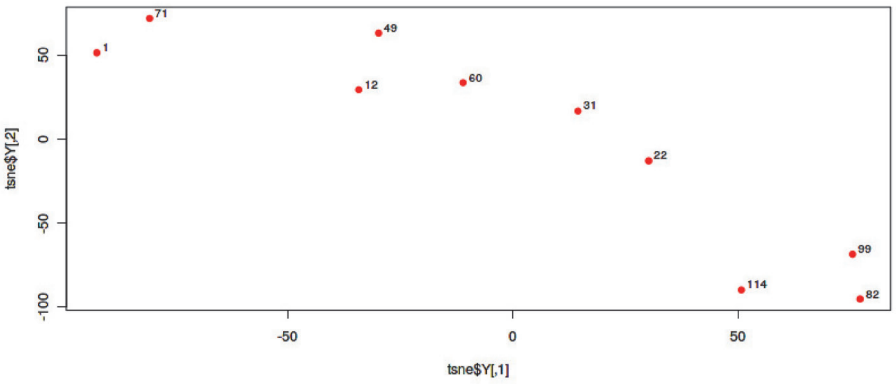

Fig.6 An example of t-SNE analysis.

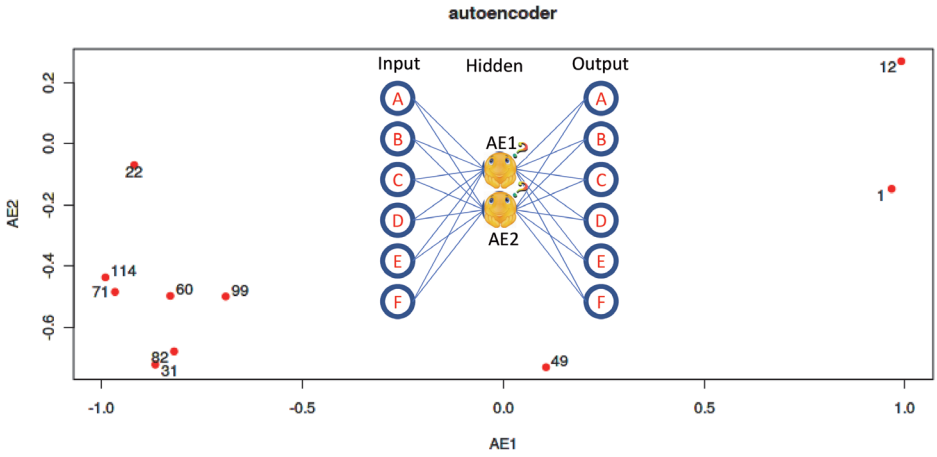

Fig.7　An example of autoencoder analysis.

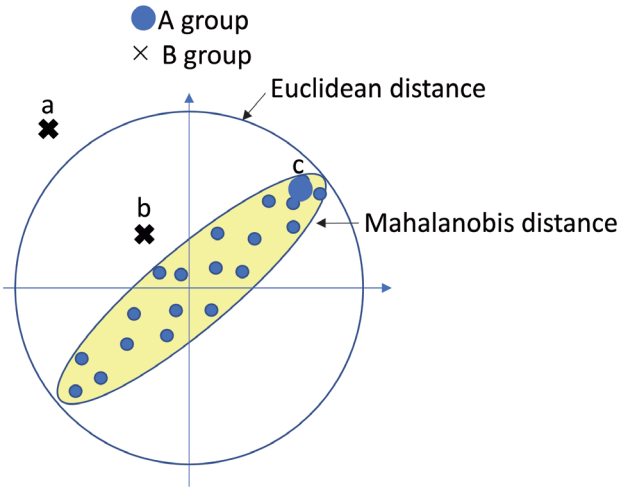

Fig.8　Euclidean and Mahalanobis distances.

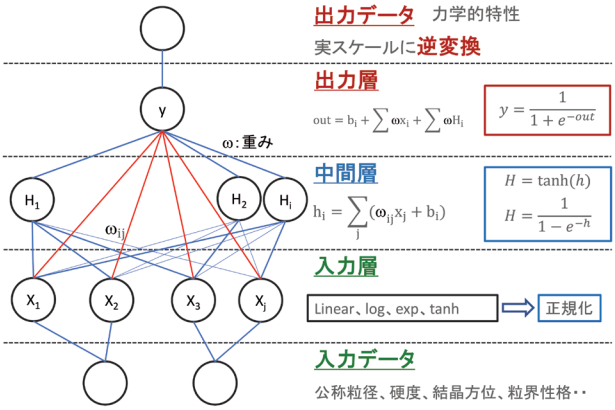

Fig.9　Architecture of neural network.

クリッド距離の指標ではAグループに所属することになってしまう。この誤分類はデータの分散方向を考慮していないことに起因する。そこで、データの分散方向を考慮した距離であるマハラノビス距離[20]（次式で与えられる）を用いると、b点はBグループに属することになり妥当である。

$$d_{i,j} = \sqrt{(x^{(i)} - m)\Sigma^{-1}(x^{(j)} - m)^T} \quad \cdots\cdots (12)$$

ここでΣ^{-1}は共分散行列、mは平均ベクトル$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$）である。

3　順解析

入力因子（例えば組織データ）と出力因子（例えば特性）間の相関を表現する近似式（モデル）を作ることが順解析である。重回帰分析では、入力層と出力層を直接直線関係で関連付け、入力層の各因子の出力層への影響度は重み係数（ω_i）で表される。しかしながらこの単純な関係式では入出力データの相関を高い精度で表現することができない場合も多い。そこで、その相関をより高い精度で表現する代表的な機械学習法（識別器ともいう）を3例紹介する。

なお、順解析モデルを作る際は、すべてのデータを正規化（平均値：0、標準偏差：1）しておくことがよい精度の順解析モデルを作るためには重要である。

3.1　ニューラルネットワーク回帰（ANN）

3.1.1　モデルの構築

ニューラルネットワーク回帰（ANN）[12]では、入力層と出力層の間に中間層（隠れ層ともいう）を設け、入力層と中間層間の関係を重み係数（ω_{ij}）を使って回帰式で表現し、さらにその回帰式を非線形な伝達関数（シグモイド関数やhyperbolic tangent関数、いずれもS字曲線を描く）で繋げることにより、一層入力層と出力層の相関を柔軟に表現していることが特徴である（Fig.9）。中間層と出力層の間は線形回

帰が用いられる場合が多い。中間層はもっとも単純なANNでは一層であるが、これを複数層にすることにより一層その精度が上がる場合がある。しかしながら、むやみに中間層を増やすことは、入力層と出力層の相関を分かりにくくすることになるため、その相関を理解できるようにした状態で回帰式を作る場合は、中間層は1層にした方がよい。

ANNモデルを作る際に全データを用いるわけではない。全データの例えば9割を使ってモデルを訓練し、そこでできたモデルに残りの1割のデータを入力しその精度を確認する(交叉検証)。たとえ訓練データを使ってできたモデルがよい精度のものであっても、交叉検証の精度が悪い場合はそのモデルは汎化されていない。このことを過学習といい、機械学習全般に共通する課題である。交叉検証は1回だけの場合もあるが、訓練データと交差検証データを入れ替えて複数回行うこともある。複数回交叉検証を行った方が、一般的により汎化されたモデルができる。

ANNはじめ機械学習の最大の課題が過学習である。過学習とは、コンピュータが訓練データに対して過度にフィッティングするモデルを構築し、そのモデルでは新規データや交叉検証用データに対して精度の良い予測ができないことである。この過学習を抑制するために、先に述べた交叉検証に加えて、以下の二つの試みが重要である。

一つ目は、入力変数全てをANNに入力するのではなく、重要な入力変数だけ(要するに記述子)に絞って入力することが重要である。特に、工業データのように、データ数に限りがあるスモールデータに対して汎化性が高いANNモデルを構築しようとする時にこの入力変数の削減は極めて重要である。先のスパース学習のところで述べた次元圧縮による入力変数も有用である。

二つ目は、過度に入力データと出力データの相関を高める(即ち実測値と予測値の誤差をできるだけ小さくする)重み係数を見つけようとするのではなく(誤差項を小さくするということと同義)、適度にその関係をスムージングしながらフィッティングするANNモデルを見つけることが重要である。この適度なスムージングとは、一つの入力因子の影響度が大きいからと言ってその重み係数を過度に大きくしすぎると過学習が生じてしまうので、一つの重み係数だけを過度に大きくしない工夫が必要であるということである。それがペナルティ項(式(13)中第二項)の導入であり、各重み係数の二乗の和と誤差項をある一定の割合(α_c:重み減衰率係数(decay))で加算したペナルティ損失関数(M(ω))を最小化するように重み係数を決定するという手法である(Fig.10)。この重み減衰率係数は式(13)で与えられる。ここでσ_ω^2は全重み係数の分散である。

$$M(\boldsymbol{\omega}) = \beta E_D + \sum_c \alpha_c E_{\omega(c)}$$
$$E_{\omega(c)}(\boldsymbol{\omega}) = \frac{1}{2}\sum_i \omega_i^2 \qquad \cdots\cdots (13)$$
$$\alpha_c = 1/\sigma_\omega^2$$

誤差項を小さくすることでフィッティングの良い相間を見つけようとするとともに、重み係数の分散をできるだけ小さくして各入力変数の影響度を同じにしようとするスムージングをバランスさせているわけである。この考えはスパース学習のところで述べたlasso回帰、ridge回帰と思想は同じである。

ANNのハイパーパラメータは中間層の層数、各中間層でのノード数(size)、重み減衰率係数である。これらのハイパーパラメータの最適化については後述する。

3.1.2 感度解析

後述するニューラルネットワーク回帰では、入力層と中間層の変数間を回帰式を非線形の伝達関数に入れて伝達し、また中間層と出力層の変数間を通常は直線回帰式で結ぶ。入力層(i)の1番目の入力変数が中間層(h)の1番目の変数に及ぼす影響度は重み係数ω_{11}と($\omega_{11}+\omega_{21}+\omega_{31}$)の比で表される(Fig.11)。同様に入力層(i)の1番目の入力変数が中間層の二番目の変数に及ぼす影響度は係数ω_{12}と($\omega_{11}+\omega_{21}+\omega_{31}$)の比で表される。

次に中間層の1番目の変数が出力層(o)の変数(m)に及

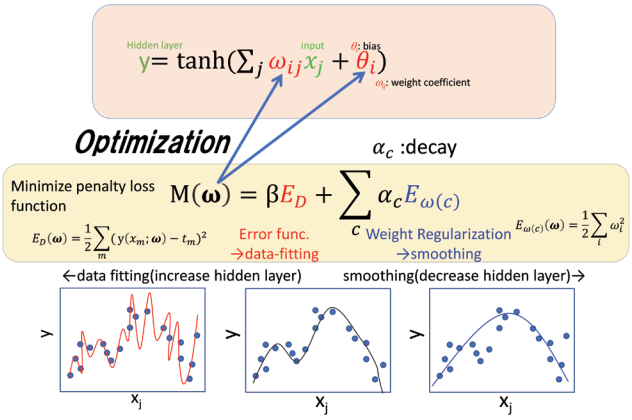

Fig.10 Penalty loss function including weight regularization to suppress over fitting in ANN.

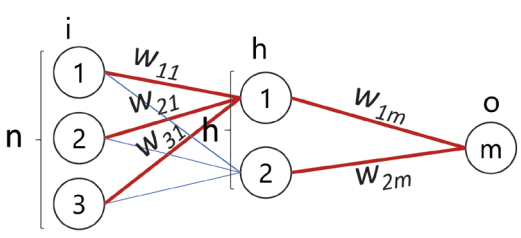

Fig.11 Sensitive analysis of ANN.

ぼす影響度はω_{1m}で表され、同様に中間層の2番目の変数が出力層（o）の変数（m）に及ぼす影響度はω_{2m}で表される。したがって、入力層の1番目の変数が出力層の変数に及ぼす影響度$t_{(1)m}$は以下のように表すことができる。

$$t_{(1)m}=\left(S_{11}\times\left|\omega_{1m}\right|\right)+\left(S_{12}\times\left|\omega_{2m}\right|\right) \quad \cdots\cdots (14)$$

同様に入力層の2番目、3番目の変数が出力層の変数に及ぼす影響を計算し比較することにより、入力変数の影響度を評価できる。このような解析を、荷重結合法による感度解析[7]という。先ほどの二相組織鋼のデータに対して解析した結果を以下に示す。

"k（1）m＝1.02537506136199"
"k（2）m＝1.42143790424327"
"k（3）m＝2.23341385619106"
"k（4）m＝0.618958978383825"
"k（5）m＝0.85772213251195"
"k（6）m＝0.625269365270278"
"k（7）m＝2.95482307202727"

7番目の入力変数であるeと、3番目のVMMと2番目のHvMが相対的に影響度が高く、それ以外は影響度が低いことが分かる。この結果は先に示したAIC, BIC, lasso回帰の結果と矛盾しない。一つの方法で変数選択することは危険であるが、複数の手法で集団学習（アンサンブル学習）することによりその判断の信頼性が増す。

感度解析には荷重結合以外に感度係数法があり、重み係数ω_i^2/分散の値を各入力変数で比較することによりその重要性を判断する。

3.1.3　ハイパーパラメータの最適化（グリッドサーチ vs. ベイズ的最適化）

すべての識別器にはハイパーパラメータがあり、分類・回帰精度をあげるためにはそれを最適化することが重要である。パイパーパラメータの最適化は従来グリッドサーチなどの全探索が行われてきたが、時間がかかるため、最近では事後確率が高くなるパラメータの領域を重点的に探索するベイズ的最適化法が注目されている。

ここでベイズ的最適化[17]の要点を説明する。例として、数点の実験結果（入力x, 出力yの事前分布）があったとき（t＝t）に、未知のブラック関数の最大値を求めることを考える。この場合評価関数は出力yそのものである。実験結果のあるところを結べばその中間点での値$\mu_t(x)$がおおよそ推定できる。この期待値が高そうなところを次は調べるべきであるという判断は正しいであろう（Fig.12）。これはいわば実験結果

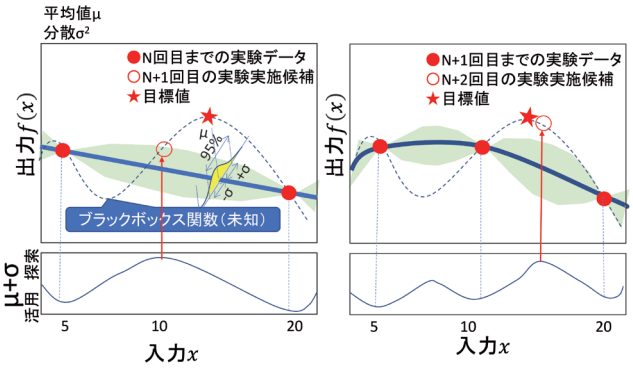

Fig.12　An explanation of Bayesian optimization.

を活用した判断である。一方で、実験結果がないところは不確定性$\sigma_t(x)$が大きく探索すべきであるという判断も正しいものと思われる。これは探索である。活用と探索の両観点を評価関数に取り入れるために、評価関数として次式を採用する。

$$\textit{評価関数}=\mu_t(x)+\kappa\sigma_t(x) \quad \cdots\cdots (15)$$

評価関数が最大値になるところが次の評価対象である。この戦略を上限信頼限界（UCB）戦略*3という。ここで活用と探索のバランスを決めている係数がκである。これをわずか数回繰り返すうちに、最大の出力値を得る入力値を見出すことができる。

この不確定性を表現するために、カーネルが用いられる。ここではよく使われるRBFカーネルを使うことにする。二点（x_i, x_j）があり、その時の出力が（y_i, y_j）とすると、例えばx_iは実験点（x_t）であり、x_jはこれから探索しようとする点（x_{t+1}）とする。x_iとx_jが近くなるほどx_i−x_jはゼロに近くなり次式のRBFカーネルのexpの値（γはここでは1とする）は1に近くなり、離れるほどその値は0に近くなる。

$$k(x_i,x_j)=\exp(-\gamma\left|x_i-x_j\right|^2) \quad \cdots\cdots (16)$$

このカーネル関数を使って、評価点（x_{t+1}）における不確定性σを次の式で得る[7]。

$$\sigma^2 t(x_{t+1})=k(x_{t+1},x_{t+1})-\boldsymbol{k}T(\boldsymbol{K}+\rho^2\boldsymbol{I})^{-1}\boldsymbol{k} \quad \cdots\cdots (17)$$

ここで、

$$\boldsymbol{k}=[k(x_{t+1},x1),\cdots,k(x_{t+1},x_t)]^T \quad \cdots\cdots (18)$$

K：式16）をij要素とするカーネルのグラム行列

$$\text{例：}K=\begin{bmatrix}k(x_1,x_1) & k(x_1,x_2) & k(x_1,x_3)\\ k(x_2,x_1) & k(x_2,x_2) & k(x_3,x_3)\\ k(x_3,x_1) & k(x_3,x_2) & k(x_3,x_3)\end{bmatrix} \quad \cdots\cdots (19)$$

＊3　その他にも、期待改善値(Expected improvement：EI)戦略、改善確率(Probability of improvement：PI)戦略などがある。

ここで、Iは単位行列、ρはデータノイズの標準偏差である。

RBFカーネルの係数γは、実験点における不確定性σの和が最小になるように決定する

$$\mu_t(x_{t+1}) = \boldsymbol{k}^T \left(\boldsymbol{K} + \rho^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y}1:\boldsymbol{t} \quad \cdots\cdots (20)$$

期待値は次式で得られる[7]。

ここで、

$$\boldsymbol{y}_{1,t} = [y_1, \cdots, y_t]^T \quad \cdots\cdots (21)$$

である。

このカーネル法によるベイズ的最適化の主旨は、x座標が似ているのであれば、y座標も似ているであろうとする考え方である。事前分布を使って、事後分布を求めていることから、逐次ベイズ的な推定手法といえる。

上述した例では評価関数が最大になるところを探索したが、二乗平均平方根誤差の場合は最小となる候補値（あるいはベクトル）を探索することになる。

一例として、ニューラルネットワークのハイパーパラメータ（sizeとdecay）の最適化をグリッドサーチとベイズ的最適化で行った結果をFig.13に示す。UCB戦略のκ値は3とした。グリッドサーチではsize（1～7）とdecay（0.05間隔で0～1）の全組み合わせ140通りをモデル式に入れて二乗平均平方根誤差を計算しており時間がかかるが、ベイズ的最適化では30通り（初期に10通りの組み合わせを与えて、その後20通りを探索）の組み合わせだけで効率よく最適値（size＝6, decay＝0.05）を割り出している。

3.2　サポートベクター回帰（SVR）

分類に用いられるサポートベクターマシン（SVM）の回帰版がSVR[13]である。プロットのマージン（d＝1/ ω）を最大化するように識別線（多次元の場合は面）が描かれるのが本手法の特徴である（Fig.14）。従って、回帰線近傍のみのデータのみを重点的に使うので、用いるデータ数が少なく一般的にANNよりも計算に要する時間が短い。回帰線を越えてもある程度は誤りを許すソフトマージン（その程度をスラック変数ξで表現する。マージンの逆数ωとξの和（そのバランスをCという係数でとっている）を評価関数としてできるだけ小さくするように回帰線を引く）という技法や、それでも回帰線が引けない場合はカーネル法により一つ上の高次元空間に座標変換して回帰線を引いたのちに元の座標に戻すことにより回帰線（あるいは超平面）を引くことが行われる。カーネルには多くの場合RBFカーネル（係数γ（σが用いられる場合もある））が用いられている。RBFカーネルについては前述した式（16）を参照願いたい。パラメータCとγ（もしくはσ）がSVRのハイパーパラメータである。Cを大きくするほど識別誤りを認めず、γを大きくするほどより複雑に回帰線を引くことになる（Fig.15）。これらのSVRの方法を採用したうえでさらに、識別線近傍の誤差をεの範囲（εチューブと呼ばれる）ではゼロにする誤差の不感帯を設けて回帰線を引くのがSVRである（Fig.14）。従って、SVRのハイパーパラ

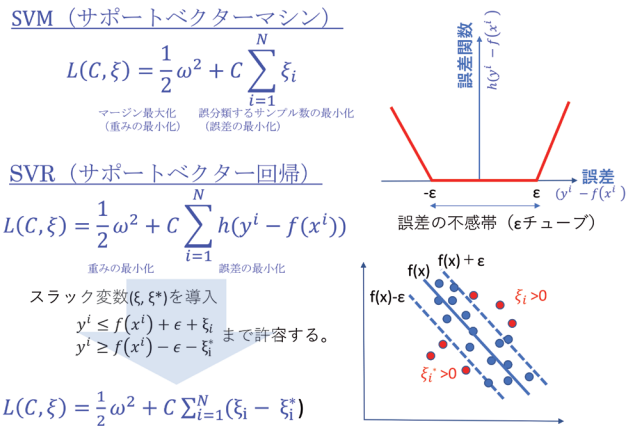

Fig.14　Comparison between support vector machine and support vector regression.

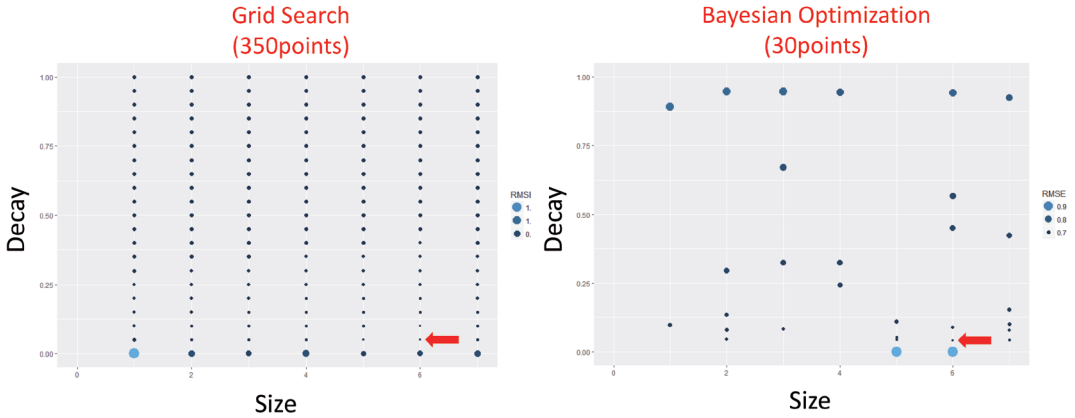

Fig.13 Hyper-parameter optimization by grid search and Bayesian optimization.

メータはC，γ（もしくはσ），εである。

3.3　ランダムフォレスト回帰（rf）

ランダムフォレスト[14]は分類、回帰に適用可能な識別器である（Fig.16）。ランダムフォレスト回帰は、複数の回帰木（決定木の回帰版）を使って、分類の場合は多数決、回帰の場合は平均値を得るアンサンブル学習である。ランダムフォレスト回帰では、全データから部分的なデータがランダムに選択（バギング）され、一回目の回帰の値によって、枝分かれ項目の中の次のどの回帰木を選択するかが決まる。回帰式の項の組み合わせもランダムに選択される。単純な方法であるがゆえに解析は一般的にANNよりも早い。

4　逆解析

材料工学における最終目標は、目的の特性あるいはトレードオフバランスの関係にある特性間の最適なバランスを有する材料を作る事である。実験、理論、シミュレーションで得たプロセス―組織―特性の結果を使ってこの目標を効率的に達成するための手法として逆解析がある。要望する特性を得るために組織あるいはプロセスを最適化するので最適化問題ともいえる。ここでいう「特性」には、単一特性に加えて、トレードオフの関係にある二つの特性のバランスを含み、二つの特性の積や和が新たな特性（目的変数）となる。

実験、理論、シミュレーションを全数探索的に行えば逆解

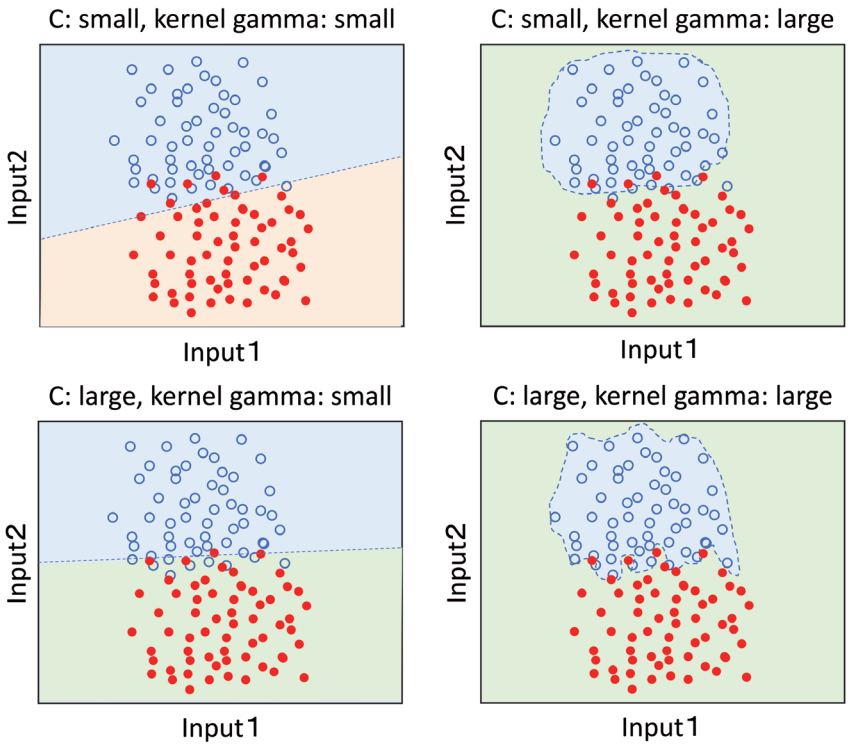

Fig.15　An effect of C and kernel g on classification/regression line in SVM/SVR.

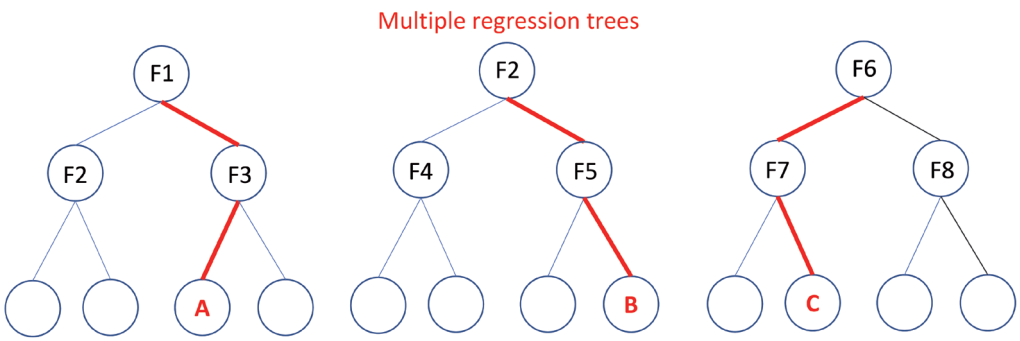

Fig.16　An explanation of random forest regression.

析ができるが、これは現実的ではない。それらの手法で得た限定的なデータを使って順解析モデルを作り、そのモデルにより「効率的に」全数探索的な計算を行うことにより逆解析を行う。以下では「効率的に」全数探索的逆解析を行う代表的な三つのアルゴリズムを紹介する。遺伝的アルゴリズム（GA）および粒子群最適化（PSO）は比較的データが多い時にデータ中の最高出力値を超える出力を得る入力変数の組み合わせを探索するのに向いているのに対して、ベイズ的最適化（BO）は数少ない実験・計算で最高出力を得たい時に有用である（実験計画法の一種である）。

4.1　遺伝的アルゴリズム（GA）

最高出力を得られるように入力変数の組み合わせを効率的に変える工夫として、GA[15]では入力変数の淘汰、再生、交叉、突然変異を一定の割合で行うことに特徴がある。これは自然界における進化論をまねたアルゴリズムである（Fig.17）。

各入力変数にある値が入っている複数のデータセットがある時に、出力が大きくなる上位一定割合のデータセットのみ生き残り、下位のデータセットは淘汰される。一部のデータセットは入力変数データの一つあるいは複数の場所を境界にデータの総入れ替えを行い新たなデータセットとする（境界の場所が決まっている場合を一点交叉といい、ランダムに境界が変わる場合を一様交叉という。また境界が複数ある場合は多点交差という）。またある一定割合のデータセットでは入力変数の値を突然変異させて新しいデータセットとする。全データの中から一世代で対象とするデータ数、最適な入力変数を求めるために繰り返す世代数もハイパーパラメータである。

Fig.18は、先に用いた二相組織材料のデータを対象に、順解析モデルはANNとし、逆解析をGAで行った結果である。ここでは、一世代あたりの訓練データ数は10、世代数は100とし、入出力データは正規化した値で表示している。正規化後の入力データは-6～6の範囲で走査した。訓練データ中の最高出力は2程度であるのに対して、GAで逆解析した結果では最高出力値は6以上となっており、訓練データ以上の出力を生む入力変数の組み合わせを見つけているということになる。なお、var1から7は、それぞれフェライトの硬度（var1＝HvF）、マルテンサイトの硬度（var2＝HvM）、マルテンサイトの体積率（var3＝VMM）、マルテンサイトの数密度（var4

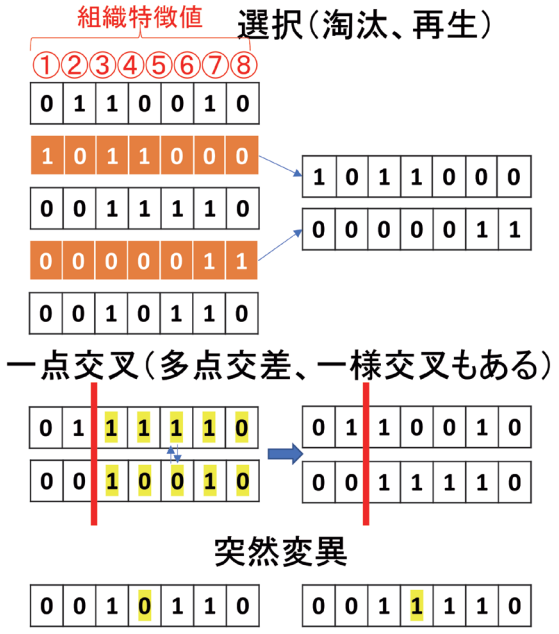

Fig.17 An explanation of genetic algorithm.

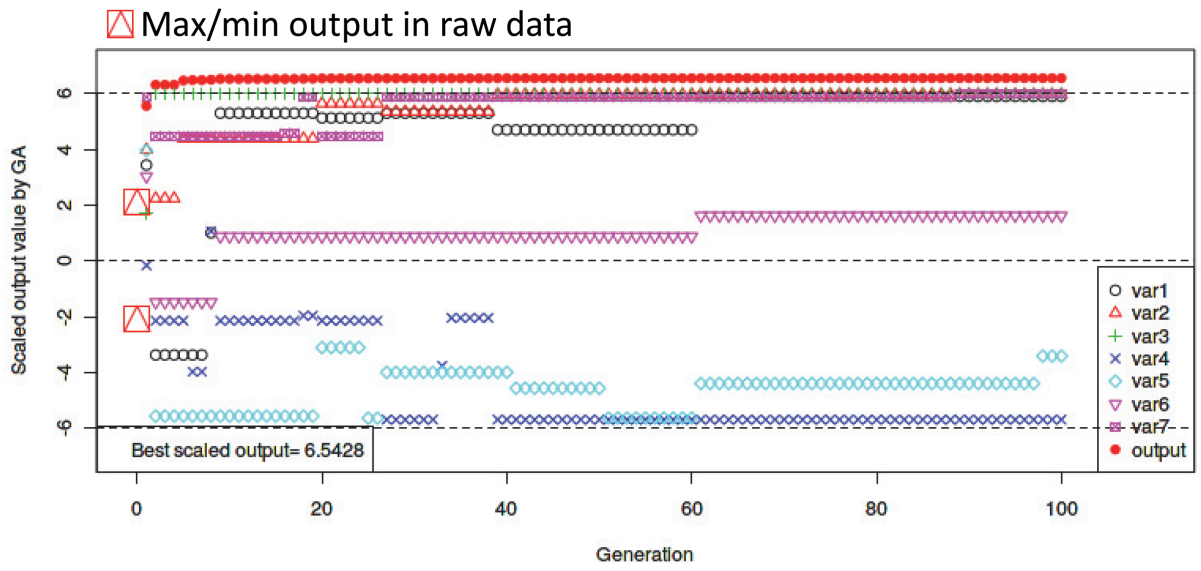

Fig.18 An inverse analysis to maximize output by ANN-GA.

＝f)、マルテンサイト中の貫通した穴の数（var5＝h)、マルテンサイト中の空洞の数（var6＝v)、そしてひずみ（var7＝e）である。outputは強度である。強度を上げるためには、HvM、VMM、eはできるだけ大きくした方がよいことを示唆している。それに対してfは小さくした方がよいという結果である。

4.2　粒子群最適化（PSO）

粒子群最適化（PSO）[16]は生物の群行動をまねたアルゴリズムである（Fig.19)。今複数の蟻が自由に動いている空間の一カ所に砂糖を置くと、一匹の蟻がたまたまその砂糖を見つけたとする。すると蟻は他の蟻に情報伝達して、他の何割（c_2）かの蟻はその方向に加速度的（vが増加）に集まって行く。その蟻が集まったところの座標が入力変数の第一次の最適値だと考える。数割（c_1）の蟻は最初の蟻の影響を受けずに自由に動き回っているので、より最適な砂糖の在処を見つけるかもしれない。したがって、局所解に陥りにくいアルゴリズムと言える。

先の二相組織鋼のデータを使って、PSO逆解析した結果をFig.20に示す。正規化後の入力データは-6～6の範囲で走査した。Fig20中には出力（強度）を最大化する時の入力変数（正規化）の値も示されている。訓練データ中の最高出力よりも高い出力が得られることをPSOは提示している。入力変数var1, 2, 3, 5, 6, 7（HvF, HvM, VMM, v, e）は平均値よりも上げた方が、一方var4（f）は下げた方が強度が上がるという結果である。この傾向は4.1で述べたGAの結果とも矛盾しない。なお、PSOはGA, BOに比べて高速である（PSO：7秒、GA：19秒、BO：23分48秒)。

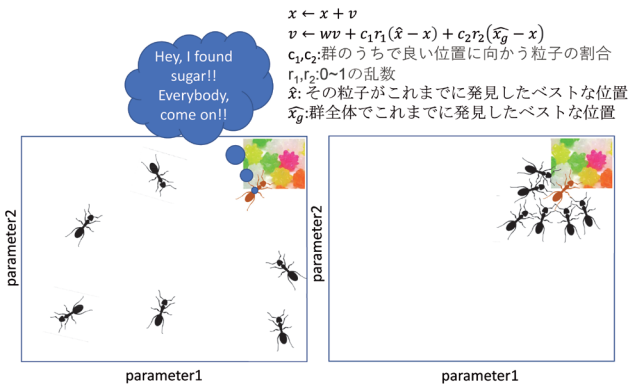

Fig.19　An explanation of particle swarm optimization.

4.3　ベイズ的最適化（BO）

3.1.3で説明したベイズ的最適化[17]は出力（通常は特性）を最大化する逆解析にも使える。先の二相組織鋼のデータを使って、BO逆解析した結果をFig.21に示す。ここでは初期学習データ数を10、学習回数を30、ハイパーパラメータのκを3と設定した。正規化後の入力データは-6～6の範囲で走査した。BO逆解析でも、訓練データ中の最高出力よりも高い出力が得られることが提示されている。特に、VMM, HvM, eはできるだけ大きくした方が、fは小さくしたほうが出力が上がるといった結果は、GA, PSO, BO共通であり信頼性が高い。なお、BOによる逆解析はGA, PSOよりも結果が安定する場合が多い一方で、計算に最も時間を要する。

5　材料情報統合システム

前報で述べた材料組織形態の定量解析および本法で述べた各種機械学習については、材料情報統合システムMIPHA[21]とshinyMIPHA[22]に実装されている（Table1)。MIPHAは画像の材料工学的に重要な特徴量抽出を得意としており、一方shinyMIPHAは数学・画像工学的に重要な特徴量抽出を得意とする。同時に、shinyMIPHAには数多くの順解析、逆解析、スパース学習モジュールが実装されている。この二つの材料情報統合システムについては本講座の続編で詳細を述べる。

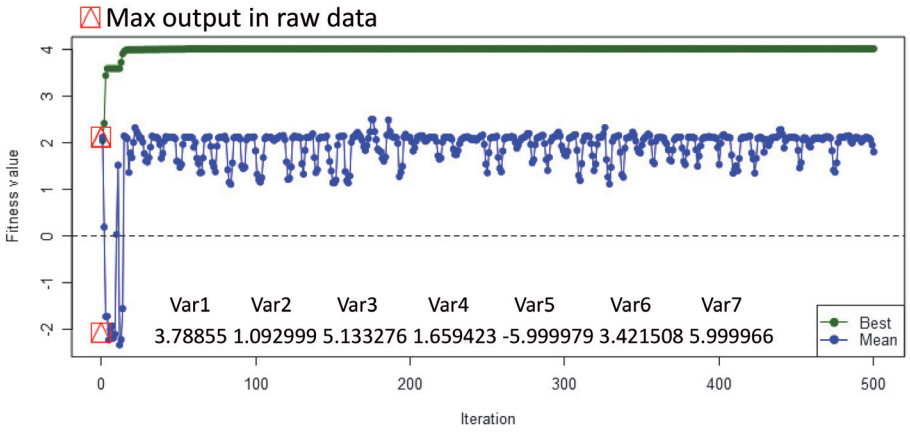

Fig.20　An inverse analysis to maximize output by ANN-PSO.

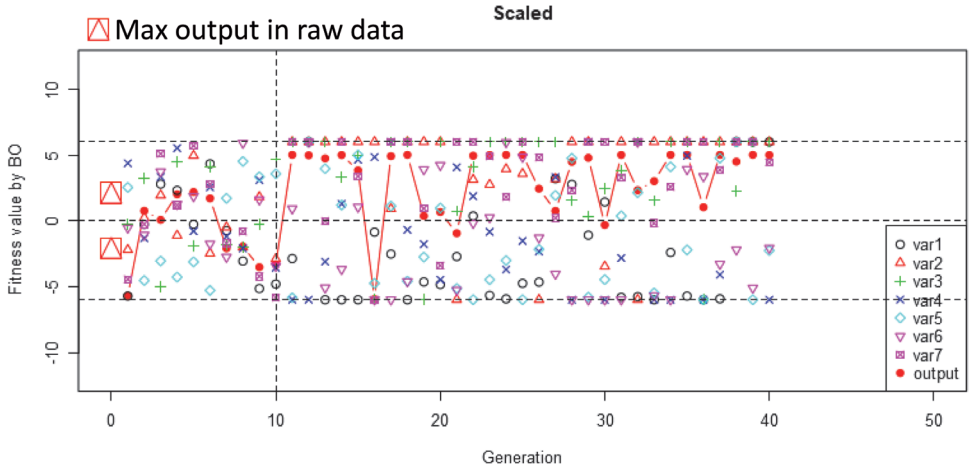

Fig.21　An inverse analysis to maximize output by ANN-BO.

Table1　Comparison in function between MIPHA and shinyMIPHA.

Content		MIPHA		ShinyMIPHA	
		2D	3D	2D	3D
Morphology	Number of particles	✔	✔	✔	
	Area/volume fraction	✔	✔	✔	
	Perimeter	✔		✔	
	Circularity/sphericity	✔	✔	✔	
	Solidity	✔			
	Grain size	✔	✔	✔	
	Ferret length	✔	✔		
	Surface area		✔		
	Gauss, mean curvature		✔		
	Genus		✔		
	Euler-Poincare		✔		
	Branching points		✔		
	Fractal dimension	✔	✔	✔	
	Morishita Iδ index			✔	
	Direction analysis			✔	
	Hough transform			✔	
	Persistent homology(PH)			✔	✔
	Inverse PH			✔	
	Two-points correlation			✔	
	Autocorrelation			✔	
	Similarity analysis			✔	
	Image processing			✔	
	Machine learning-based image processing	✔	✔		
Sparse modelling	AIC/BIC/lasso			✔	
	Glasso			✔	
	PCA			✔	
	Kernel PCA			✔	
	t-SNE			✔	
	Autoencoder			✔	
Clustering	K-means			✔	
	SOM			✔	

Direct analysis	Regression analysis		✔
	Generalized linear mixed-effects model		✔
	Gaussian process reg.		✔
	ANN	✔	✔
	SVR		✔
	Recurrent NN		✔
Inverse analysis	ANN-genetic algorism	✔	✔
	ANN-particle swarm opt.		✔
	ANN-Bayesian opt.		✔
	SVR-Bayesian opt.		✔
	RF-Bayesian opt.		✔
Else	Create random data		✔
	Create dataset		✔
	Repeat image		✔
	Select variables		✔
	Random sampling		✔
	Calculate mean/std		✔
	Bind CSV file		✔
	2D/3D plot		✔
	Histogram		✔
Microstructure Analysis*	Fast Fourie Transform		✔
	Stereographic projection		✔
	Deviation from OR		✔
	Schmid factor		✔
	Modified JMAK eq.		✔
	Histogram peak fitting		✔
Mec. *	SS curve by swift eq.		✔
	Micromechanics		✔

*Implemented in shinyMIPHA_plus

6 材料工学への適用が期待されるその他の機械学習

AlphaGo[23] に象徴される多くの教師データを初期に与えなくても自ら適応していくアルゴリズムである強化学習 (reinforcement learning) [24] も、プロセス条件の最適解を求めるアルゴリズムとして魅力的である。

初期値から未来の値を予測する Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) [25] は、クリープや疲労などの時系列データの推定に有用であると推察されるがまだ適用されたという報告例はない。

Generative Adversarial Network (GAN) [26] は、実物 (例：画

像）に似ているように偽物（例：疑似画像）を作るgeneratorと実物か偽物化を識別するdiscriminatorの二つが切磋琢磨し、教師なし学習をして実物に極めて類似する偽物を作る手法である。自立型画像処理とも言え、実像の周辺の類似画像を増やしたい時などに使えるが、材料工学への適用はまだ進んでいないが今後の展開を注視したい技術である。

7 まとめ

入力変数（組織特徴量など）と出力変数（特性など）の数値データが揃えば、その相関を機械学習によりモデル化することができる。両者の関係が非線形で複雑であっても、機械学習は究極的な近似式を作ることによってその相関を表現する。理論、シミュレーションでは両者の関係は因果関係である必要があるが、相関関係も最大限に使う機械学習がその代替になるわけではない点は注意が必要である。あくまで機械学習は限られた数の実験結果、理論式による計算結果、シミュレーションで得られた結果を総合的にまとめ上げる際に有用なツールである。機械学習は、一般的に実験や、シミュレーションに比べて、計算が早いので、その全数探索的計算を行えば逆解析が実質的に可能である点もその特徴の一つである。

参考文献

1）足立吉隆, Z.L.Wang, 小川登志男：ふぇらむ, 25（2020）9, 569.
2）足立吉隆, Z.L.Wang, 小川登志男：ふぇらむ, 25（2020）10, 628.
3）例えば, 冨岡亮太：スパース性に基づく機械学習, 講談社サイエンティフィック, 東京,（2015）
4）H.Akaike：Proc. of the 2nd International Symposium on Information Theory, ed. by B.N.Petrov and F.Caski, Akadimiai Kiado, Budapest,（1973）, 267.
5）H.S.Bhat and N.Kumar：On the derivation of the Bayesian information criterion, School of Natural Sciences, University of California,（2010）
6）R.Tibshirani：J. R. Stat. Soc. B, 58（1996）, 267.
7）齊藤進, 加藤雅啓, 坂本知子：八戸工業高等専門学校紀要, 37（2002）, 49.
8）H.Hotelling：J. Educ. Psychol., 24（1993）, 417.
9）G.E.Hinton and R.R.Salakhutdinov：Science, 313（2006）5786, 504.
10）B.Schölkopf, A.Smola and K.-R.Müller：Proc. of International Conference on Artificial Neural Networks,（2005）, 583.
11）C.E.Rasmussen and C.K.I.Williams：Gaussian Processes for Machine Learning, The MIT Press,（2016）
12）R.J.Schalkoff：Artificial neural networks, McGraw-Hill, New York,（1997）
13）M.A.Hearst, S.T.Dumais, E.Osuna, J.Platt and B.Scholkopf：IEEE Intell. Syst. App., 13（1998）, 18.
14）A.Liaw and W.Matthew：R news, 2（2002）, 18.
15）M.Mitchell：An Introduction to Genetic Algorithms, The MIT Press, Cambridge,（1998）
16）R.C.Eberhart and Y.H.Shi：Proc. of the 2001 Congress on Evolutionary Computation, IEEE, Piscataway, NJ,（2001）, 81.
17）M.Pelikan, D.E.Goldberg and E.Cantú-Paz：Proc. of the 1st Annual Conf. on Genetic and Evolutionary Computation, Vol. 1, Morgan Kaufmann Publishers Inc., Burlington, MA,（1999）, 525.
18）足立吉隆, Z.L.Wang：ふぇらむ, 23（2018）12, 672.
19）L.van der Maaten and G.Hinton：J. Machine Learning Research, 9（2008）, 2579.
20）P.C.Mahalanobis：Proc. of the National Institute of Sciences of India, 2（1936）1, 49. 2（1）：49.
21）Z.L.Wang and Y.Adachi：Mater. Sci. Eng. A, 744（2019）28, 661.
22）Z.L.Wang, T.Ogawa and Y. Adachi：Journal of Advanced Theory and Simulations, 1900177（2019）, 1.
23）D.Silver, A.Huang, C.J.Maddison, A.Guez, L.Sifre, G.van den Driessche, J.Schrittwieser, I.Antonoglou, V.Panneershelvam, M.Lanctot, S.Dieleman, D.Grewe, J.Nham, N.Kalchbrenner, I.Sutskever, T.Lillicrap, M.Leach, K.Kavukcuoglu, T.Graepel and D.Hassabis：Nature, 529（2016）, 484.
24）R.S.Sutton and A.G.Barto：Reinforcement Learning：An Introduction（Adaptive Computation and Machine Learning）second edition, ed. by F.Bach, A Brandford Book, The MIT Press, Cambridge, Massachusetts, London, England,（1998）
25）X.Liang, L.Lin, X.Shen, J.Feng, S.Yan and E.P.Xing：arXiv：1703. 03055v1,（2017）
26）I.J.Goodfellow, J.Pouget-Abadie, M.Mirza, B.Xu, D.Warde-Farley, S.Ozair, A.Courville and Y.Bengio：arXiv:1406.2661v1,（2014）
27）Z.L.Wang, T.Ogawa and Y.Adachi：ISIJ Int., 59（2019）9, 1691.

（2020年1月7日受付）