材料工学における数学，情報工学への更なる期待

足　立　吉　隆[*1)]　Zhi-Lei Wang[*2)]

1. 緒　　言

材料開発には時間を要する「作業」があまりにも多すぎる．無用な絨毯爆撃的実験，組織写真中の対象領域の手動色塗り，結晶粒径・体積率などの手動評価，複雑系での実験結果の傾向把握などは時間を要する作業となっており，それらを効率化することにより研究者，技術者はクリエイティブな思考に時間をより一層割くことができるようになるであろう．機械学習は，これらの効率化，推定，解析のいずれの点でも活用でき，研究開発の生産性向上に向けて積極的に導入されるべきである．機械学習をその特徴から大別(**図1**)すると，教師あり学習，教師なし学習，強化学習に分けることができる．以下では，すでに使われている方法もあれば，まだ利用されていない方法もあるが，材料工学への適用への期待を込めて所感を述べたい．なお，本稿では，物質探索ではなく，材料開発における数学および情報工学の適用に関する著者の考えを述べさせていただきたい．

2. 推　　定

複雑系におけるプロセス→組織→特性の関係を表現するする精度の良い順解析モデル(direct analysis)を，背後にあるメカニズムの解明と並行して，構築したい．そして，この順解析モデルを使って逆解析(inverse analysis)し，特性→組織→プロセスの提案を行って材料開発を効率的に加速したい．このような場面には機械学習の**教師あり学習**が有用である．

順解析モデルは二つに大別[(1)]され，入力―出力の関係を表現する理論式のパラメータ(係数やべき指数)を線形関数で推定するパラメトリック法，非線形の関数を組み合わせて入力―出力の関係を究極的な近似式で表現するノンパラメトリック法がある．単純系でデータノイズが小さい場合には物理的背景がある理論式の精度を向上することにつながるパラメトリック法が優先して用いられるべきであり，一方複雑系でデータノイズが大きい場合にはノンパラメトリック法がパラメトリック法よりも精度が優れる場合がある．ノンパラメトリック法はパラメータで式を表現しない究極的な近似モデルといえ，その利用をためらう意見もあるが，最終手段として活用し，解析手法の候補から排除するべきではないというのが著者の主張である．

パラメトリック法は，最尤法でパラメータを推定する一般線形モデル General linear model(lm)，一般化線形モデル Generalized linear model(glm)と，事後確率を最大化する方法でパラメータを推定するマルコフ連鎖モンテカルロ法(MCMC)がある[(1)]．データは通常ノイズを含んでおり，ポアソン分布，ガウス(正規)分布，ガンマ分布などの確率密度

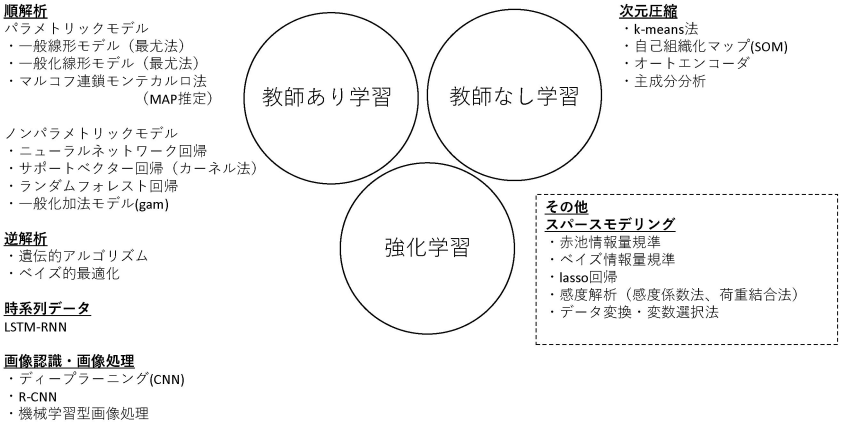

図1　機械学習の3大分野．

* 名古屋大学大学院工学研究科材料デザイン工学専攻：1)教授　2)研究員(〒464-8603 名古屋市千種区不老町)
Further Expectation of Mathematics and Information Engineering in Material Science and Engineering; Yoshitaka Adachi and Zhi-Lei Wang(Nagoya University, Graduate School of Engineering, Department of Material Design Innovation Engineering, Nagoya)
Keywords: *materials informatics, descriptor, parametric model, non-parametric model, inverse analysis, sparse modelling, deep learning*
2018年8月21日受理[doi:10.2320/materia.58.29]

関数と，尤度や事後確率などの統計学に関する基礎知識が必要である．ノンパラメトリックな手法としては，最も簡単なのは，一般化加法モデル（Generalized additive model (gam)）であり，これは非線形関数でプロットをつなげて曲線を描く手法である．そのため，明示的に数式では表現できないが，線形モデルよりも一般的に精度がよい．さらに入力層と出力層の間に中間層（隠れ層という場合もある）を入れて表現力を増したのが，ニューラルネットワーク（artificial neural network: ANN）であり，機械学習の代表的な手法と言えるであろう．その他にも非線形関係を扱う機械学習法として，カーネル法を取り入れたサポートベクター回帰（support vector regression: SVM），複数の決定木（decision tree）の多数決で回帰を行うランダムフォレスト（random forest: RF）などの手法がある．ANN, SVM, RF などを識別機（classifier）と呼ぶ．いずれの識別機にもモデル構造の詳細を表すハイパーパラメータ（例えば，ニューラルネットワークの場合には中間層中のノードの数や重み減衰率係数など）があり，精度の良い順解析モデルを作るためには，その最適化が重要である．ハイパーパラメータの最適化には，ベイズ的最適化やグリッドサーチが有用である．

精度の良い順解析モデルができると，そのモデルを使って逆解析が可能となる．逆解析時に，最高出力となる条件を探索する際の入力変数の範囲は，通常順解析モデルを作った時の範囲であるが，出力は初期データの範囲外となる場合もある．最適条件の探索アルゴリズムとしては，グリッドサーチ（全探索），遺伝的アルゴリズム[2]，ベイズ的最適化[3]とあり，後者に行くほど数少ない探索で最高出力条件を見つける効率的な探索を行う．

AlphaGo[4]に象徴される，教師データを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）[5]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．材料工学への適用はまだこれからといったところである．強化学習では最初に多くのデータを与えるのではなく，一回の試行ごとに報酬あるいは罰を与えて，動作を良い方向に導く手法である．

初期値から未来の値を予測する Long Short-Term Memory Recurrent Neural Network（LSTM-RNN）[6]は，文脈の解析に用いられるという説明が多いようであるが，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．しかしながら，まだ適用例はなく今後の進展が期待される．

3. 解　　析

準備する入力変数（記述子ではない）は多いほど良い．しかしそのまま機械学習に入力するのではなく，次に述べる交互作用項を設けたり，重要な入力変数を選択したり，入力変数の次元を削減したりすることが重要である．

2 つの記述子の相互作用が出力に影響を大いに与えている場合は，記述子は単独で与えるだけではなく，記述子同士の交互作用項を作って入力変数（記述子候補）とする場合がある[7]．

入力変数を削減し重要な記述子（交互作用項を含む）のみを機械学習に入力することは過学習を抑制する上で極めて重要であり，スパース学習† の重要な部分といえるであろう．この記述子の重要性を機械学習で判断する手法として，赤池情報量基準（AIC）[8]，ベイズ情報量基準（BIC）[9]，lasso 回帰，感度係数法[10]や荷重結合法[11]による感度解析，データ変換・変数選択法[12]などがある．この入力変数を選択する方法とともに，次に述べる次元削減法も入力変数を削減する有効な方法である．

k-means 法（**図 2**）[13]は，与えられた入力データだけを使って分類対象の集合を部分集合に分割する（クラスタリングという）ことが可能であり，主成分分析（principle component analysis），オートエンコーダー（auto encoder）などと同様に**教師なし学習**の一手法である．これらの手法は多次元情報の次元削減手法ともいえる．k-means 法の計算手順は以下のとおりである．1. 各データ x_i（$i=1...$ n）に対してランダムに k 個のクラスにクラス分けする（クラスタリング），2. 割り振ったデータをもとに各クラスタの中心 V_j（$j=1...$ k（平均）を計算する，3. 各 x_i と各 V_j との距離を求め，x_i を最も近い中心のクラスタに割り当て直す，4. 上記の処理で全ての x_i のクラスタの割り当てが変化しなかった場合，あるいは変化量が事前に設定した一定の閾値を下回った場合に，収束したと判断して処理を終了する．そうでない場合は新しく割り振られたクラスタから V_j を再計算して上記の処理を繰り返す．

k-means 法は代表点に属する近くのプロットを集めて部分集合化するだけの手法であったが，特徴量を有した状態でクラスタリングする手法としては自己組織化マップ（Self-organizing map: SOM，**図 3**）[14]があり，これはニューラルネットワークの教師なし学習である．SOM の内容は以下のとおりである．

1. 入力データ $x_{i,n}$ とニューロン $m_{i,j,n}$ 間の距離 d を測定し，最小距離のニューロンを勝者ニューロンとする．

$$min(d)=\|\boldsymbol{X_i}-\boldsymbol{M}_{i,j}\|=\sum_{n=1}^{n}(x_{i,n}-m_{i,j,n})^2$$

（例）$\boldsymbol{M}_{3,6}=(m_{3,6,1}, m_{3,6,2}, m_{3,6,n})$*を勝者ニューロンとする．*

2. 勝者ニューロンを次式に従って修正する．

$$M_{3,6}^{new}=M_{3,6}+\alpha(X_i-M_{3,6})$$

α：学習率（$0<\alpha<1$）

3. 近くのニューロン（例 $M_{2,6}$）を修正する．

$$M_{2,6}^{new}=M_{2,6}+\alpha(X_i-M_{2,6})\times g(e)$$

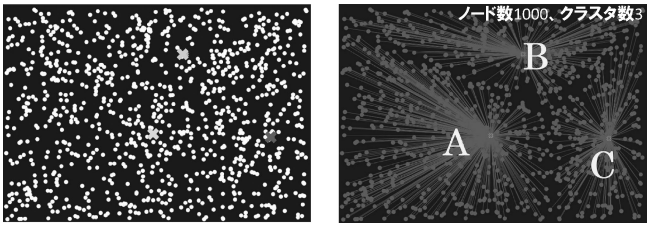

図 2　k-means 法．

† スパース学習は，データの本質的に意味のある情報の低次元性を利用し，目的に関係ない情報を削除しながら学習をする方法である．

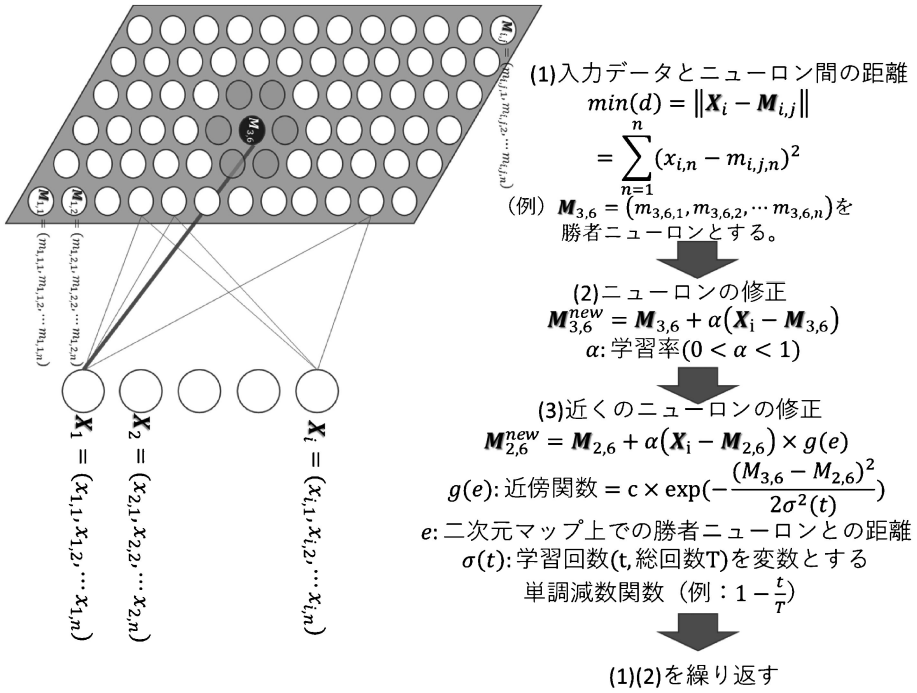

図 3　自己組織化マップ.

$$g(e)：近傍関数 = \mathrm{c} \times \exp\left(-\frac{(M_{3,6} - M_{2,6})^2}{2\sigma^2(t)}\right)$$

e：二次元マップ上での勝者ニューロンとの距離

$\sigma(t)$：学習回数(t, 総回数 T)を変数とする

$$単調減数関数\left(例：1 - \frac{t}{T}\right)$$

1〜3 を繰り返すことにより，位相情報が似ている入力データは自己組織化マップでは徐々に近くに集まってくるようになる.

4. 効　率　化

異なる視点として，分析機器のユーザーフレンドリー化に機械学習法を導入することが期待される．画像もスペクトラムもその特徴量を数値化して誰でも可読な形式(例えば csv 形式)で保存されることが望まれる．そしてその定量解析がワンクリックでできるように機械学習を活用することが考えられる．特徴量の定量評価には数学が活用でき，位相幾何学，微分幾何学，さらにはフーリエ変換法の発展形で階層的に潜在する連結性を評価するパーシステントホモロジー群などがすでに材料工学に適用され始めている[15)–17)]．材料工学の専門家に与えられた使命は，それぞれの分野において重要な記述子(descriptor)を設計することである．繰り返しになるが，これらの豊富な特徴量を，容易に得られるようにすることが極めて重要である.

材料組織の画像処理では，手作業で対処領域に色を塗る(segmentation あるいは labeling という)ことに多くの時間を費やしている研究者も多いものと推察される．この作業については，深層学習(deep learning)に代表される機械学習型画像認識・画像処理は飛躍的な進展を続けており[18)19)]，材料工学においてもっと活用されるべきと思われる.

5. 材料情報統合システム

以上述べた数学手法や機械学習法を容易に材料研究者が使いこなせるような材料情報統合システムの構築が真に材料開発効率を革新的に改善するために重要である．幸い各種機械学習法のモジュールがプログラミング言語 Python や R で提供されており，材料研究者はそれらのモジュールを自由自在に組み合わせて自らの課題に適用させることが求められる．したがって今後の材料研究者はこれらのプログラミング言語の習得が不可欠になるであろう．教育現場では，これらのプログラミング言語を自由自在に使いこなせる人材を育成するためのカリキュラムの抜本的改革が期待される.

著者たちは，上述した機械学習法の大半を取り組んだ材料情報統合システム Material genome integration system for phase and property analysis(MIPHA† および rMIPHA_studio††)(**図 4**)を開発し，提供を始めている[20)–22)]．同時に，「統合型材料デザイン」という大学院生向けの講義を開設している.

† MIPHA：プログラミング言語 visual basic，画像認識，機械学習型画像処理，定量 2D/3D 解析，ニューラルネットワークによる順解析，遺伝的アルゴリズムによる逆解析が可能である.

†† rMIPHA_studio：プログラミング言語 R， rMIPHA， Theory designer, Material image editor の三つのモジュールで構成される．rMIPHA は ANN, SVR, RF による順解析，ベイズ的最適化によるそれぞれの識別機を使った逆解析，スパース学習が実装されており，Theory designer は理論式のパラメトリック推定が実装されており，Material image editor は画像間の類似性評価が実装されている.

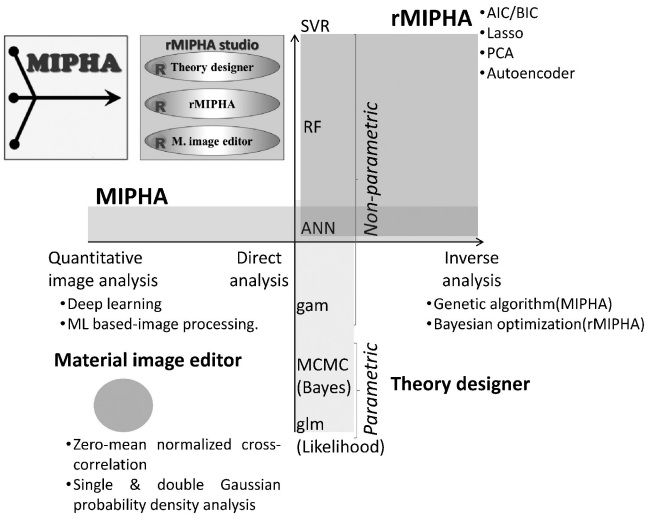

図４　材料情報統合システム．

6.　その他材料工学への展開が期待される情報工学

Generative Adversarial Network(GAN)[22]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，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.　結　　言

人工知能(AI)に代表される情報工学の発展は目覚ましいが，一昔前の AI でもいいので材料工学へ適用するという意識改革がいま求められている．材料工学の課題を情報工学研究者に丸投げするのではなく，材料工学研究者が自ら AI を学び材料工学に適用しなければ材料開発の真の革新的効率化は頓挫するであろう．これまでは数学，物理，化学を材料工学に適用することが進められてきたが，今後はそれらに加えて AI を適用することが当然という時代がすでに到来している．それぞれの分野の専門家が記述子を設計し，従来型の理論式による予測を行うとともに，AI(機械学習)を使った究極的な近似式を使った推定を並行して行うことができる人材の育成が急務である．

最後に，AI を活用した材料工学(マテリアルズインフォマティクスと最近では言われているが，情報工学主体ではなくあくまで材料工学主体であるべきという主張を込めてインフォメイティブマテリアルサイエンスと呼ぶべきと著者は考えている)は，今後材料工学を担っていく若手の材料研究者，技術者に開かれた新たな研究分野であり，興味を持つ若手研究者，技術者が多く参画されることを大いに期待したい．

文　　献

（１）C. M. Bishop 著，元田　浩，栗田多喜男，樋口知之，松本祐治，村田　昇　監訳：パターン認識と機械学習 上，丸善出版，(2012)．
（２）J. R. Koza: Genetic Programming, MIT Press, Massachusetts, (1993)．
（３）E., Brochu, et. al: arXiv:1012.2599v1.
（４）D Silver, et. al: Nature, **529** (2016). 484–489.
（５）R. S. Sutton (Author), A. G. Barto (Author), F. Bach (Editor), Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning) second edition, A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England, (1998).
（６）X. Liang, L. Lin, X. Shen, J. Feng, S. Yan and E. P. Xing: arXiv:1703.03055v1, (2017).
（７）例えば，青木　敏，武村彰通：確率・統計Ⅱ，丸善出版，(2018), 22.
（８）H. Akaike: Proceedings of the 2nd International Symposium on Information Theory, B. N. Petrov, and F. Caski. (eds.), Akadimiai Kiado, Budapest, (1973), 267.
（９）文献(1), 215.
(10) 弘川奨悟，田口茂樹，松下康弘，足立吉隆：鉄と鋼，DP 鋼の応力―ひずみ曲線を支配する組織因子のスパース学習，鉄と鋼，**103**，(2017). 468–474.
(11) S. Saito, M. Kato and T. Sakamoto: Bulletin of Toyama Technical College, No. 37, (2002), 49.
(12) J. W. Tukey: Exploratory Data Analysis, Addison–Wesley, Indianapolis, (1977).
(13) J. B. MacQueen: Proc. of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, California, (1967), 281.
(14) T. Kohonen: Biological Cybernetics, **43**(1982), 59–69.
(15) 足立吉隆，小山敏幸：3D 材料組織・特性解析の基礎と応用―シリアルセクショニング実験およびフェーズフィールド法からのアプローチ―，内田老鶴圃，(2014)．
(16) Y. T. Wang, Y. Adachi, K. Nakajima and Y. Sugimoto: ISIJ International, **52**(2012), 697–703.
(17) 平岡裕章(著)，三村昌泰(編集)，竹内康博(編集)，森田善久(編集)，タンパク質構造とトポロジー パーシステントホモロジー群入門(シリーズ・現象を解明する数学)，共立出版，2013–07–11.
(18) 足立吉隆，田口茂樹，弘川奨悟：鉄と鋼，**102**(2016), A 722–729.
(19) 田口茂樹，弘川奨悟，安田　格，徳田耕平，足立吉隆：先進的画像処理による組織検出，鉄と鋼，**103**(2017), 142–148.
(20) 足立吉隆，松下康弘，上村逸郎，井上純哉：機械学習支援の材料情報統合システム，システム制御情報学会誌，**61**(2017), 188–193.
(21) Z.–L. Wang and Y. Adachi: Mater. Sci. and Eng. A, July, 2018. (Under review).
(22) Z.–L. Wang and Y. Adachi: Advanced Materials, August, 2018. (Under review)
(23) I. J. Goodfellow, J.–Abadie, M. Mirza, B. Xu, D. Warde–Farley, S. Ozair, A. Courville and Y. Bengio: Generative Adversarial Networks, (2014).

足立吉隆

★★★★★★★★★★★★★★★★★★★★★★
1990年３月　名古屋大学大学院工学研究科修士課程修了
主な略歴　物質・材料研究機構，鹿児島大学
2017年４月–現職
専門分野：高次元材料情報学
◎鉄鋼材料を主対象に材料組織を定量 3D 解析し，そこで得た材料ゲノム情報を使って順解析・逆解析モデルを構築する研究に従事している．
★★★★★★★★★★★★★★★★★★★★★★