表面科学における計測インフォマティクス ～歴史的展開と将来展望～

藤 田 大 介*

国立研究開発法人 物質・材料研究機構 ☎ 305-0047 茨城県つくば市千現 1-2-1

(2018 年 12 月 26 日受付；2019 年 1 月 29 日掲載決定)

Measurement Informatics in Surface Science ~Historical Evolution and Future Prospect~

Daisuke FUJITA*

National Institute for Materials Science, 1-2-1, Sengen, Tsukuba, Ibaraki 305-0047, Japan

(Received December 26, 2018 ; Accepted January 29, 2019)

Following the historical development of informatics, data-intensive science, machine learning and materials informatics, we introduce important applications to surface science. As an indication of first application of multivariate analysis to surface science, application of target factor analysis to depth profiling data by Auger electron spectroscopy is introduced including the detailed analysis principle. We believe the way the future measurement system should be is a “smart measurement system” linked with cyber space and physical space. In such a data driven measurement system, novel informatics that can be applied to big measurement data are required. From such a viewpoint, the fusion of surface analysis and informatics, that is, “surface measurement informatics” is expected as a new paradigm of surface science.

KEYWORDS : informatics, data intensive science, machine learning, chemometrics, multivariate analysis

1. インフォマティクスとデータ集約型科学

データ科学を駆使した新しいタイプの材料研究，すなわち“マテリアルズ・インフォマティクス”が脚光を浴びている。本稿ではインフォマティクスならびに関連する学術領域，ならびに計測分析分野への適用である計測インフォマティクスの歴史的発展を辿りつつ，最近の表面科学への応用展開を紹介し，今後 10 年程度の近未来展望についても概観したい。そもそも“インフォマティクス (informatics)”という科学用語は，1957 年にドイツの K. Steinbuch (1917~2005) によりドイツ語として造語された“informatik”に起源すると思われる¹⁾。“自動情報処理 (automatic information processing)”の副題が論文には付随するが，informatik もしくはその英語に対応す

る informatics の範囲はより広く，いわゆる“計算機科学” (computer science) として当時は理解されていた。情報と計算の理論的基礎及びその計算機上への実装と応用に関する研究分野である。なお，Steinbuch は人工ニューラルネットワーク (artificial neural network) のパイオニアでもあり，ドイツにおける計算機科学の草創期において影響力のある研究者であった。1990 年代に英国エジンバラ大学にインフォマティクス学部が設置されたが，そこではインフォマティクスはコンピュータサイエンスよりも広く定義されている²⁾。すなわち，自然または人工のシステムにおける情報の表現 (representation)，処理 (processing)，伝達 (communication) について研究する学問領域とされる。コンピュータ，個人および組織はすべて情報を処理するため，インフォマティクスは計算的，認知のおよび社会的な様相を有する。まさに情報の科学を示している。

* E-mail : fujita.daisuke@nims.go.jp

Fig. 1. (color online). Schematic research cycle of data-intensive science to totally visualize a data iceberg.

ところで、データも情報の一形態であり、インフォマティクスにはデータ科学 (data science) も包含される。近年、我が国においても人口に膾炙し始めた“データ科学”の学術用語は、データ集約型科学 (data-intensive science) もしくはデータ駆動型科学 (data-driven science) とも呼ばれる。多様な形式のデータから知識や洞察を抽出する科学的プロセスやシステムに関する学際領域である。データ科学は、統計学、機械学習、データマイニング、予測解析などのデータ分析分野の継続ともみなされるし、データベースによる知識発見 (Knowledge Discovery on Database, KDD) とも類似する。筆者の考えるデータ科学は、①様々な計測手法を用いたデータ取得 (data capture) から始まり、②データの組織化と統合を図るデータキュレーション (data curation) を経て、③情報科学的な手法でデータ解析 (data analysis) を行い、水面に現れている情報のみならず水面下に埋もれた情報をも抽出することにより、④冰山として擬物化されたデータ (data iceberg) の全体を可視化 (data visualization) するリサーチサイクルを構成する (Fig. 1)。

計算機科学におけるノーベル賞とも云われるチューリング賞 (Turing award) 受賞者である J. Gray (米国、マイクロソフト研究所) は、2007 年 1 月に米国研究評議会 (National Research Council, NRC) のコンピュータサイエンス & 電気通信委員会 (Computer Science and Telecommunications Board, CSTB) において、“eScience”に関する重要な講演を行った。eScience とは、高度に分散したネットワーク環境、またはグリッドコンピューティングを必要とする膨大なデータセットを使用する計算集約型の科学である。そこにおいて Gray は、人類の科学の発展史を 4 つの類型に分類した (Fig. 2)。まず、数千年前からの経験的な科学 (empirical science) に始まり、数百年前から数式に基づく理論科学 (theoretical science) を経て、20 世紀後半からコンピュータを活用した計算

Fig. 2. (color online). Historical evolution of the paradigms of science.

科学 (computational science) へとパラダイムシフトが起きてきたことを示した。引き続いて、今日ではデータ集約型科学 (data-intensive science) の時代にあり、科学における「第 4 のパラダイム (fourth paradigm)」であるとした³⁾。

2. マテリアルズ・インフォマティクス

既存の学術領域の名称とインフォマティクスが複合した場合、複合された研究分野に特化した情報学、情報科学を意味している。メディカル・インフォマティクスやバイオ・インフォマティクスなどが該当する。例えば、バイオ・インフォマティクス (生物情報学、生命情報科学) は、生物学 (biology) や生命科学 (life science) のデータを理解するための解析法やソフトウェアを情報科学の手法に基づいて開発する学際的な領域である。1980 年代後期から、“bioinformatics”の用語は主としてゲノムデータの比較分析のための計算手法や解析手法を指すために用いられた。その後、近年のゲノム解析による生じた大量のデータの蓄積 (ビッグデータ) と効率的な解析手法の必要性により急速に発展してきた。しかしながら、1970 年頃に Hogeweg らにより造語された“bioinformatics”は、“生命システムの情報学的なプロセスの研究”と当初より広く定義されていた^{4,5)}。最近では、トレンドとしては、バイオ・インフォマティクスの“用語の原義”が再び明示的に現れているように思われる。

マテリアルズ・インフォマティクス (materials informatics) はおそらくバイオ・インフォマティクスの

成功に触発された用語であると考えられる。インフォマティクスを材料科学に応用した研究領域である。従来、新材料の発見や合成から産業への応用展開、すなわち実用化にいたるまで凡そ20年以上かかっていた（と言われている）。材料の科学や工学にインフォマティクスの手法を応用することにより、材料の発見、開発、選択、使用について加速することを目的とする学際分野である。

マテリアルズ・インフォマティクスは21世紀初め頃からRajan（米国）らより提唱されたものと思われる。2005年のMaterials Today誌に掲載されたRajan単著論文のタイトルは“Materials Informatics”そのものである⁶⁾。構造と機能の関係を探索することは材料科学のパラダイムの一つであるが、これらの関係はしばしば線形でない。マルチスケール時空間で関係性を探索することがチャレンジとなる。これまで単一の理論や実験でこのような関係性の情報を抽出することは困難であった。Rajanはそのような複雑かつマルチスケールの関係性を、高いスループット、統計的に安定、かつ物理的に有意に抽出できる手法、プロセスとして“Materials Informatics”（材料情報学）を提案した。このようなアプローチは、材料の設計と発見において重要なインパクトがあると指摘している。データベースによる知識発見（KDD）、言い換えるとデータマイニング（data mining）により、大規模材料データベースに相関と理論を融合させることにより、新材料発見につながる知識ベースを抽出することが目的である^{7,8)}。

$$\text{Data} + \text{Correlations} + \text{Theory} = \text{Knowledge-base} \quad (1)$$

我が国においてもマテリアルズ・インフォマティクスの重要性への認識は2005年頃から十分にされていた。物質・材料研究機構は2005年において2020年を見据えた将来の物質・材料研究のあるべき姿を議論し、「2020年NIMSポリシーペーパー」として発表した⁹⁾。そのレポートにおいて「ポストナノテクノロジー・イニシアティブ」として、近未来におけるマテリアルズ・インフォマティクスの重要性を以下の様に指摘した。“多数の物質・材料を一括合成し、高速評価、データ処理を統合した集積化コンセプトに基づく物質・材料研究手法は、少子高齢化社会においても競争力を維持・向上させるうえで、グローバルな戦略性を有している。さらに、ものづくりの電子システム化に伴って発生する膨大なマテリアルデータを高速かつ有効に処理するマテリアルインフォマティクスと呼ばれる新しい研究分野も派生している。”

3. 計測インフォマティクス

計測分析分野におけるインフォマティクス応用の嚆矢は、おそらくは計量化学、すなわちケモメトリクス（chemometrics）にまで遡ることができる。ケモメトリクスは多変量解析などのデータ駆動型手法により、化学システムから得られる大量の計測データから有用な情報を抽出する学際分野を指す。1970年代前半にかけて、Kowalski（米国）、Wold（スウェーデン）らにより提唱されたが、“chemometrics”の用語は、Woldによって初めて用いられた¹⁰⁾。多くの初期のアプリケーションは多変量解析の手法を駆使したものである。1970年代後半から始まった計測分析機器のコンピュータ制御と計測データストレージの急速な進歩により、ファクターアナリシスなどの多変量データの解析手法が発展したものである¹¹⁾。

これらの多変量解析の手法の多くは、マテリアルズ・インフォマティクスにおけるデータマイニングの機能と重なるものがある。データマイニングは、統計学（statistics）、機械学習（machine learning）、人工知能（artificial intelligence）、パターン認識（pattern recognition）などの融合した学際領域である¹²⁾。機械学習とデータマイニングで用いられる手法には重畳する部分が多いが、次のように目指す目的により区別される。まず、機械学習は文字通り、機械もしくはコンピュータによる学習もしくは訓練を実現する技術分野であり、その目的の一つは、訓練データから学んだ「既知」の特徴に基づく予測である。従来の機械（計算機）はプログラムにより規定されたルールに従って答えを出すのが、機械学習は、そのルールを大量のデータから機械に学ばせる。一方、データマイニングの目的は、それまで「未知」だったデータの特徴を発見することである。大量なデータの中で埋もれていた情報を抽出することができる。データマイニングのプロセスは、通常は多段階のプロセスで構成されており、回帰分析（regression）、クラスターリング（clustering）、次元削減（dimensionality reduction）などがある。例えば、主成分分析（principal component analysis, PCA）は多変量解析の最も基本的な手法であり、データの線形性を利用して、元データを高次元データ空間から低次元データ空間に変換し、データ空間の次元削減に用いられる。計測分析分野においてはPCAなどの次元削減をしつつ埋もれた情報を抽出する多変量解析手法が1980年代頃から欧米を中心として応用されるようになった。

ところで、今日に至るまでの情報通信分野の進展は、1950年代後半の集積回路（Integrated Circuit, IC）の発明に起因して開始された大規模集積回路（Large Scale

Fig. 3. (color online). Historical evolution of informatics and related data sciences along with Moore's law.

Intergradation, LSI) の発展による質 (計算速度) と量 (メモリー, ストレージ) の著しい増大により表現することができる。単位チップあたりのトランジスタ集積度増大プロセスがムーアの法則に従っており, 18~24ヶ月で倍増することから, この法則によれば, 半導体の性能は指数関数的に向上する。この情報通信分野における質的量的な発展過程はインターネットの普及と相俟って, グローバルなスケールで生成される情報量の指数関数的な増加, すなわち情報爆発 (information explosion) をもたらす。この観点からムーアの法則と並べてインフォマティクスの発展に関連するマイルストーンを年表にした (Fig. 3)。現在のマテリアルズ・インフォマティクスの基盤となる提案や学際領域は 1950 年代から 1980 年代までに形成されていることがわかる。マテリアルズ・インフォマティクスが世界的な潮流となったのは 2011 年の米国マテリアルズ・ゲノムイニシアティブ (Materials Genome Initiative, MGI) 以降である。MGI は新材料の発見と設計に要する研究開発時間を半減することにより米国製造業を発展させることを目的とする。

4. 表面スペクトルデータ応用の嚆矢

1980 年代になると Gaarenstorom らにより, 表面化学分析手法のスペクトルデータへの PCA の応用が報告され始めたが 1990 年代初め頃までは大規模データへの適用事例はほとんどみられなかった^{13, 14)}。2000 年以降は PCA を中心として, 様々な表面分析手法への多変量解析の応用事例が増加している^{15, 16)}。我が国における表面分析分野の比較的大きなスペクトルデータへのインフォマティクス応用の嚆矢として, 1991 年から筆者らが開始したオージェ電子分光法 (Auger Electron Spectroscopy, AES) による深さ方向分析 (depth profiling) データへのターゲットファクターアナリシス (target factor analysis)

の応用事例について解析原理を含めて詳しく紹介する¹⁷⁾。

表層の組成や状態の深さ方向分布を求める手法としては, 表面分析手法である AES と希ガス (Ar^+) イオンビームスパッタリングを組合せた深さ方向解析 (depth profiling) が一般的である。この場合, 化学種 AES 強度としてスペクトルの数個の点強度を利用して得られるピーク高さを用いるのが一般的であった。広く普及している円筒鏡型電子分光器 (Cylindrical Mirror Analyzer, CMA) を用いた AES 装置の場合, 得られるスペクトルは $EN(E)$ もしくはその微分型 dEN/dE である。ここで E は電子運動エネルギー, $N(E)$ は放出電子エネルギー分布を示す。 $EN(E)$ 型の場合, ピーク最大点とバックグラウンドを代表する点との差を利用する。より普及している微分型 dEN/dE スペクトルの場合, 正側の極大点と負側の極小点の差 (Auger Peak-to-Peak Height, APPH) が用いられる。

しかしながら, ピーク高さ測定による解析はピーク重量がある場合や同一元素の状態変化には機能しない。化学種を代表するスペクトルはある程度の幅のエネルギー領域にわたっているのに対し, ピーク高さ測定はその領域内のわずかな点からの情報しか利用していないことによる。当時, 多変量解析手法の一つであるファクターアナリシス (factor analysis, 因子分析) が化学分野で得られるスペクトル解析に応用され成功を取め始めていた。ファクターアナリシスとは, “多数のスペクトルデータを固有値解析することによって, 存在する化学種の数と濃度を抽出” する手法である。界面における濃度プロファイルの形状解析は従来 APPH 測定によって行われていたが, データ駆動型計測においては多変量ビッグデータに対応するインフォマティクスの手法が利用されるようになる。

4.1 スペクトルの加成性

ファクターアナリシスなどの多変量解析手法の多くはスペクトルの加成性 (additivity) を前提条件としている。加成性は, 試料 j においてエネルギー i におけるスペクトル強度 d_{ij} が化学種 k の濃度 c_{kj} の線形結合 (linear combination) で表現できることを意味する。

$$d_{ij} = \sum_k^n r_{ik} c_{kj} \quad (2)$$

ここで, r_{ik} は化学種 k が単位濃度である場合の標準スペクトル中のエネルギー i におけるスペクトル強度を表わす。 n は試料中に存在する化学種の数である。

事例として, Si(100) 基板上に成膜された TiN/Ti 二層薄膜の AES デプスプロファイル測定時の低運動エネルギー領域における一次微分オージェスペクトル (144

Fig. 4. (color online). First-derivative 144 Auger spectra obtained for TiN(100 nm)/Ti(100 nm)/Si(100) during AES depth profiling.¹⁸⁾ E and N stand for kinetic energy and energy distribution of emitted electrons.

本)を **Fig. 4** に示す¹⁸⁾。主要化学種である TiN, Ti, Si に対応するスペクトルは多数回出現するため比較的高い線として表われている。Fig. 4 で 1 から 6 までの点は TiN/Ti 界面もしくは Ti/Si 界面において観測された不動点である。不動点は各標準スペクトル間の交点であり、加成性が成立すれば界面近傍の二相混合スペクトルにおいても位置が不変であるべき点である。不動点の存在からスペクトルの加成性の成立が保証される。

次に (2) を行列を用いて表現すると以下の様になる。

$$[D] = [R][C] \quad (3)$$

データ行列 $[D]$ は各列ベクトルが各測定スペクトル (スペクトルデータ点数 r , 測定スペクトル数 c) により構成されている ($r \times c$) の行列である (**Fig. 5**)。

オージェ電子分光スペクトル計測から得られた一次データには前処理 (pretreatment) としてはノイズ除去のための Savitzky-Golay 二次三次多項式適合平滑化処理やバックグラウンド除去のための微分処理を必要に応じて施される¹⁸⁾。特に加成性がないバックグラウンドの効果を低減するために一次微分 (dEN/dE), もしくは二次微分 (d^2EN/dE^2) の各タイプのスペクトルを用意できる。 $[R]$ はロー行列 (row matrix) と呼ばれ, 存在する n 個の化学種の標準スペクトルを各列ベクトルにもつ ($r \times n$) の行列である。 $[C]$ はコラム行列 (column matrix) と呼ばれ, 各化学種の測定スペクトル (1 ~ c) 中に占める濃度を各行ベクトルにもつ ($n \times c$) 行列である。

4.2 多次元データの次元縮減

ファクターアナリシス解析とは, データ行列 $[D]$ から出発して数学的処理によってデータ空間の次元数 n すなわち構成化学種数や物理的に意味のある $[R]$ と $[C]$ を求

Fig. 5. (color online). Mathematical relation among data matrix $[D]$, row matrix $[R]$ and column matrix $[C]$.

めることである。最初の数学的処理は, $[D]$ にその転置行列 $[D]^T$ を乗じて, ($c \times c$) の共分散行列 $[Z]$ を作ることである。

$$[Z] = [D]^T[D] \quad (4)$$

共分散行列 $[Z]$ は正方行列かつ対称行列であり, その固有ベクトルはデータ行列 $[D]$ のファクター空間を張るベクトルと同一である。 $[Z]$ を対角化することによって固有値行列 $[\lambda]$ と固有ベクトル行列 $[Q]$ を求める。

$$[Z][Q] = [\lambda][Q] \quad (5)$$

$[\lambda]$ の対角成分 λ_i は各固有値 ($\lambda_1 \geq \lambda_2 \geq \dots$) であり非対角成分は 0 である。 $[Q]$ の各列ベクトルは各固有値に対応した固有ベクトルである。各固有ベクトルはファクター空間を張る基底単位ベクトルであり正規直交系を成している。ゆえに

$$[Q]^{-1} = [Q]^T \quad (6)$$

真の次元数は n であるが, この段階ではファクター空間は誤差の効果も含むために c 次元もしくは r 次元の多次元空間である。いくつかの経験的方法 (IE 関数, IND 関数, variance など) によって誤差範囲内で $[D]$ を再構成するのに必要十分な主成分数 n に次元を削減することができる。各固有値は各固有ベクトルの重要度を定量的に表している。また, IE 関数 (imbedded error function) と IND 関数 (factor indicator function) は適切な次元数を決めるために Malinowski により提案されており, 真の次元数 n 値で極小となる¹¹⁾。variance 値は固有ベクトルの相対重要度を示すものである。事例として, GaAs/AlAs 超格子系におけるデプスプロファイルの低運動エネルギー領域一次微分オージェスペクトル (150 本) から成るデータ行列 $[D]$ の共分散行列 $[Z]$ の固有値解析から得られた全固有値の常用対数表示を **Fig. 6** に示す¹⁹⁾。

左端の丸 (点線) で囲まれた固有値 2 個 (λ_1, λ_2) が他の固有値より二桁以上大きいことから直感的に $n=2$ が予測される。また, Fig. 6 に挿入された IE 関数がフ

Fig. 6. (color online). Plot of eigenvalues obtained for the 150 AES spectra of GaAs/AlAs superlattice depth profiling. Insert is the plot of Imbedded Error function.¹⁹⁾

ファクター数が2の値で極小をとることからも真の次元数 n が2であると判定できる。

主成分の数 n がわかることによって $[A]$ および $[Q]$ も n 次元の $[A^+]$ と $[Q^+]$ へ変換される。上記の数学的プロセスから式 (3) を満足する解の1組である $[R^+]$ と $[C^+]$ が得られる。これらは抽象解 (abstract factor) と呼ばれる。

$$[C^+] = [Q^+]^T \quad (7)$$

$$[R^+] = [D][Q^+] \quad (8)$$

$$[D^+] = [R^+][C^+] \quad (9)$$

この段階で得られた抽象解を物理的に意味のある解にするにはファクター空間の基底を変換行列 $[T]$ によって回転 (変換) させなければならない。変換行列を求めるためにターゲットテスト等のプロセスを行う。これらのプロセスから各化学種の深さ方向濃度変化プロファイルが得られる。このように、ファクターアナリシス法は初期においては多層膜系深さ方向解析データ (数百本程度のスペクトルデータ) における次元削減と成分数決定や界面プロファイルの定量評価に効果を発揮した (Fig. 7)。また、GaAs/AlAs 超格子系のような原子レベルで急峻なヘテロ界面を有する超薄膜試料の深さ方向解析における“深さ分解能を規定する要因”の定量的な抽出を可能にした²⁰⁾。さらに、SiO₂/Si 界面に存在すると考えられていた SiO_x 成分の AES 深さ方向解析データからの抽出や電子ビーム照射による還元状態の抽出にも効果を発揮している²¹⁾。

4.3 測定データのノイズ縮減

実際のスペクトル測定ではノイズ等に起因する測定誤差が必ず存在するため、実測されたデータ行列 $[D]$ は以

Fig. 7. (color online). Comparison of AES depth profiles of GaAs/AlAs superlattice obtained by Al KLL APPH measurements and by Al LMM factor analysis.¹⁹⁾

Fig. 8. (color online). Relation of real error (RE), extracted error (XE) and imbedded error (IE).

下の様に表現できる。

$$[D] = [D^*] + [E] \quad (10)$$

ここで $[D^*]$ は純粋データ行列 (pure data matrix), $[E]$ は誤差行列 (error matrix) である。ファクターアナリシスにより次元削減によって再構成されたデータ行列 (reproduced data matrix) $[D^+]$ は抽象解を用いて式 (9) で表される。Malinowski は次元削減により抽出可能なノイズ (誤差) を XE (extracted error), 除去できずに埋め込まれるノイズ (誤差) を IE (imbedded error) とし、実際の測定ノイズ (誤差) を RE (real error) との関係を Fig. 8 の様に直角三角形で表現できることを示した¹¹⁾。ここで IE は RE と次元数 n を用いて以下の様に表される。

$$IE = RE \sqrt{\frac{n}{c}} \quad (11)$$

測定スペクトル数 c が大きいほど残留誤差である IE を小さくでき、抽出誤差 XE を実測誤差 RE に近づけるこ

とができることがわかる。これはビッグなスペクトルデータの固有値解析と次元縮減により計測ノイズを大幅に除去し、従来は検出不可能であった埋もれた情報も可視化できることを示している。

5. 表面計測インフォマティクスの応用展開

表面分析手法への最近の応用展開の事例を紹介したい。表面分析の基礎基盤となる分野へのファクターアナリシス法の応用が活発化している。表面電子分光法において基礎物理パラメータとなる非弾性散乱平均自由行程 (*Inelastic Mean Free Path*, IMFP) 導出に必須の誘電関数やエネルギー損失関数を実験的に反射電子エネルギー損失分光 (*reflection electron energy loss spectroscopy*) のスペクトルデータ群から抽出する試みが Tanuma らによって行われている^{22~24)}。また、表面分析データベースの構築は、データ科学の本格的活用には不可欠であるため、共通データ転送フォーマットの提案が Yoshikawa らにより行われている²⁵⁾。

表面の物性や構造を原子分解能で計測可能な走査型プローブ顕微鏡 (*Scanning Probe Microscopy*, SPM) は多次元スペクトルイメージングに適した手法である。力学的、物理的、化学的特性と表面形状を多次元同時測定する SPM 分光イメージングデータはテラバイトへ向けてビッグデータ化してきている²⁶⁾。SPM の中でも表面電子状態を原子レベルで解明可能な走査型トンネル分光 (*Scanning Tunneling Spectroscopy*, STS) においては、表面再構成やドーパント、点欠陥などが多様な電子状態を形成しており、複数の電子状態が重畳したスペクトルデータから存在状態数と対応する位置情報を抽出する手法の開発がデータ科学の導入によって加速されるものと考えられる²⁷⁾。さらに共焦点ラマン顕微鏡や ToF-SIMS 等の 3 次元スペクトルデータセットへの適用も進展している²⁸⁾。

6. 近未来の展望

インフォマティクスを表面分析などの先端的な計測分析手法へ導入し、データ科学と融合した計測の新たなパラダイムを構築することが期待されている。従来の計測機器はスタンドアロンでオペレータが操作し、インターネット等を駆使した外部制御や外部への自動データ転送等の機能は想定されていなかった。また、データ解析はエキスパートの経験に負うことが多かった。将来の計測システムのあるべき姿を想定すると、Fig. 9 に示すようにサイバー空間と物理空間と連結した“スマート計測システム”が必須となるだろう。制御系には人工知能 (*Artificial Intelligence*, AI) を導入することより、オペレ

Fig. 9. (color online). Schematic representation of data-driven smart measurement system.

ータを介せずに計測条件の最適化とハイスループットの測定を行うことができる。物理空間の計測システムは外部サイバー空間と高速通信網で接続され、外部からの制御 (*input*)、外部データベースへの自動データ出力 (*output*)、データ科学・計算科学クラウドによるデータ解析、ソリューションクラウドによる最適な計測制御条件の提供等のイノベーションが実現される。

このようなデータ駆動型計測システムには多次元ビッグスペクトルデータに適したインフォマティクス手法が必要不可欠とされるだろう。そのような観点から表面計測とインフォマティクスの融合、すなわち“表面計測インフォマティクス”が近未来の表面分析パラダイムとして期待される所以である。

文 献

- 1) K. Steinbuch : “Informatik : Automatische Informationsverarbeitung (Informatics : automatic information processing)” (SEG-Nachrichten, Berlin, 1957).
- 2) M. Fourman : “informatics”, Informatics Research Report EDI-INF-RR-0139, Division of Informatics, University of Edinburgh, U.K. (2002).
- 3) T. Hey, S. Tansley and K. Tolle : “The Fourth Paradigm : Data-Intensive Scientific Discovery” (Microsoft Research, 2009).

- 4) B. Hesper and P. Hogeweg : *Kameleon* **1**, 28 (1970). (In Dutch.).
- 5) P. Hogeweg : *PLOS Comput. Biol.* **7**, e1002021 (2011).
- 6) K. Rajan : *Mater. Today* **8**, 38 (2005).
- 7) K. Rajan : “Informatics for Materials Science and Engineering, Data-driven Discovery for Accelerated Experimentation and Application” (Elsevier, 2013).
- 8) T. Lookman, F.J. Alexander and K. Rajan : “Information Science for Materials Discovery and Design” (Springer, 2016).
- 9) 物質・材料研究機構 : “2020 年 NIMS ポリシーペーパー” (2006).
- 10) S. Wold : *Chemom. Intell. Lab. Syst.* **30**, 109 (1995).
- 11) E.R. Malinowski and D.G. Howery : “Factor Analysis in Chemistry” (Wiley, 1980).
- 12) C.M. Bishop : “Pattern Recognition and Machine Learning” (Springer, 2006).
- 13) S.W. Gaarenstroom : *Appl. Surf. Sci.* **7**, 7 (1981).
- 14) V. Atzrodt and H. Laxge : *Phys. Status Solidi A* **79**, 489 (1983).
- 15) J.C. Vickerman and I. Gilmore : “Surface Analysis, The principal Techniques” (Wiley, 2009).
- 16) K.M. Mc Evoy, M.J. Genet and C.C. Dupont-Gillain : *Anal. Chem.* **80**, 7226 (2008).
- 17) 藤田大介, 吉原一紘 : *表面科学* **13**, 286 (1992).
- 18) A. Savitzky and M.J.E. Golay : *Anal. Chem.* **36**, 1627 (1964).
- 19) 藤田大介, 吉原一紘 : *表面科学* **14**, 324 (1993).
- 20) K. Yoshihara, D.W. Moon, D. Fujita, K.J. Kim and K. Kajiwarra : *Surf. Interface Anal.* **20**, 1061 (1993).
- 21) D. Fujita, K. Onishi, T. Yakabe and K. Yoshihara : *J. Surf. Anal.* **13**, 190 (2006).
- 22) H. Jin, H. Yoshikawa, S. Tanuma and S. Tougaard : *Surf. Interface Anal.* **42**, 1076 (2010).
- 23) H. Jin, H. Shinotsuka, H. Yoshikawa, H. Iwai, S. Tanuma and S. Tougaard : *J. Appl. Phys.* **107**, 083709 (2010).
- 24) H. Jin, H. Shinotsuka, H. Yoshikawa, H. Iwai, M. Arai, S. Tanuma and S. Tougaard : *Surf. Interface Anal.* **45**, 985 (2013).
- 25) H. Yoshikawa, K. Yoshihara, D. Watanabe, H. Tanaka and S. Tanuma : *Surf. Interface Anal.* **46**, 931 (2014).
- 26) A. Belianinov, S. Kalinin and S. Jesse : *Nat. Commun.* **6**, 6550 (2015).
- 27) J. Yamanishi, S. Iwase, N. Ishida and D. Fujita : *Appl. Surf. Sci.* **428**, 186 (2018).
- 28) H.-X. Wang, H. Zhang, B. Da, M. Shiga, H. Kitazawa and D. Fujita : *J. Phys. Chem. C* **122**, 7187 (2018).