

解説

材料情報の多面的利用

吉武 道子*

物質・材料研究機構

〒305-0003 茨城県つくば市桜3-13

*Yoshitake.michiko@nims.go.jp

(2018年1月28日受理; 2018年3月26日掲載決定)

近年、コンピュータ技術の発達によって大量のデータを統計的に解析して有用な情報を取り出す動き、ビッグデータ解析、が劇的に加速している。材料分野でもここ2~3年ぐらい、盛んに取りざたされるようになった。ただ、この手法は大量のデータが存在してこそ威力を発揮する手法で、材料分野においては、大量の数値データが存在しないという問題点を抱えていると著者は感じている。本稿では、そのままコンピュータに入力できる数値データのデータ量不足を補う手法として、著者が取り組んでいる内容を報告する。

Utilization of Information on Materials in Multi-Dimensional

Michiko Yoshitake*

National Institute for Materials Science

3-13, Sakura, Tsukuba, Ibaraki 305-0003 Japan

*Yoshitake.michiko@nims.go.jp

(Received: January 28, 2018; Accepted: March 26, 2018)

Recent development of computer technology has made possible to analyze huge datasets statistically and extract practically useful information, which is called 'big-data analysis'. In materials science field, this type of approach has been in fashion in these two or three years. However, this approach is powerful with huge datasets, which are not fully supplied in materials science. The author has developed a few techniques to complement the shortage of numerical data that can be input directly to the computer, which are briefly reported in this article.

1. はじめに

物質・材料分野において、物質・材料のデータに対し情報技術を駆使することによって材料探索しようという、マテリアルズインフォマティクスが最近盛んになっている[1]。ただ、実験によりデータを取得するのが必ずしも容易では無いため、情報処理に用いることのできるデータ量がかなり限られる。そのため現在は主に、第一原理計算で電子状態や基本的な物性値を求めて、それをデータとして用いた研究が大勢を占める。

一方、材料情報は、誘電率やXPSスペクトルのよ

うな計測されたデータやシミュレーション(計算)データだけではなく、材料の作製条件など計測・計算データのメタデータとみなされるデータや、科学法則のような知識データも含まれる。

計測・計算データやメタデータと計測・計算データの両方を用いた解析は、コンピュータに機械学習させる方法が盛んである。この手法は基本的には、従来「多変量解析」と呼ばれた手法およびその延長線上にある手法で、ある範囲内の(組成、作製方法など)データが多量に存在する際に、非常に効率よく目的変数の最適化(最適組成や最適アニール温度

など)が可能である。自動車の交通量や顧客の商品購買履歴など、日々大量に生み出される情報から、交通量を予測したり商品のお薦めを提示するなど、いわゆるビッグデータ解析の世界である。

しかし、科学法則のような知識データは、単純に数値列に置き換えられないこともあり、活用はまだ十分とは言えない。また、計測・計算データについても、発表されている論文の数は膨大であるにも関わらず、コンピュータによる機械学習に向けた数値データ化されたデータで共有化できるものはほとんど無い。

このような状況下、論文から数値化されたデータを作成する方向と、データが少なくても科学原理(知識データ)を利用して材料探索を可能にする方向の、大きく分けて二通りのアプローチから研究開発を行っている。本稿では、その概要 (Fig. 1) を紹介する。

2. グラフ (スペクトル) 画像自動数値化ソフト

Fig. 1 の「論文から数値化されたデータを作成す

る方向」として、論文中のグラフ (スペクトル) 画像を自動で数値化するソフトを開発している。本項ではこのソフトについて述べる。開発した自動数値化ソフトは、論文や紙ベースのグラフ (スペクトル) スキャン PDF ファイルから、あらかじめ数値化したグラフを bmp 形式で抜き出して (複数)、特定のフォルダーに格納しておいてから用いる。

各グラフ画像の横軸・縦軸の最小・最大値を自動で読み込んで、フォルダーに存在している画像を次から次へと自動的に処理し、変換した数値データを CSV 形式のファイルとして出力する。処理できるグラフのタイプは、Fig. 2 に模式的に示したように、(a) 一本の直線グラフ、(b) 多段グラフの分割、(c) 重なりのない複数グラフ、(d) 破線・点線グラフ、(e) 多色の重なりのあるグラフ、である。なお、横軸・縦軸の最小・最大値を自動検出するために、Windows 上でのみ動作する無料 OCR ソフトを使用している。その関係上、この自動数値化ソフトは Windows 上でのみ動作する。また、本自動化ソフトは NIMS 著作物登録されており、NIMS 外部連携

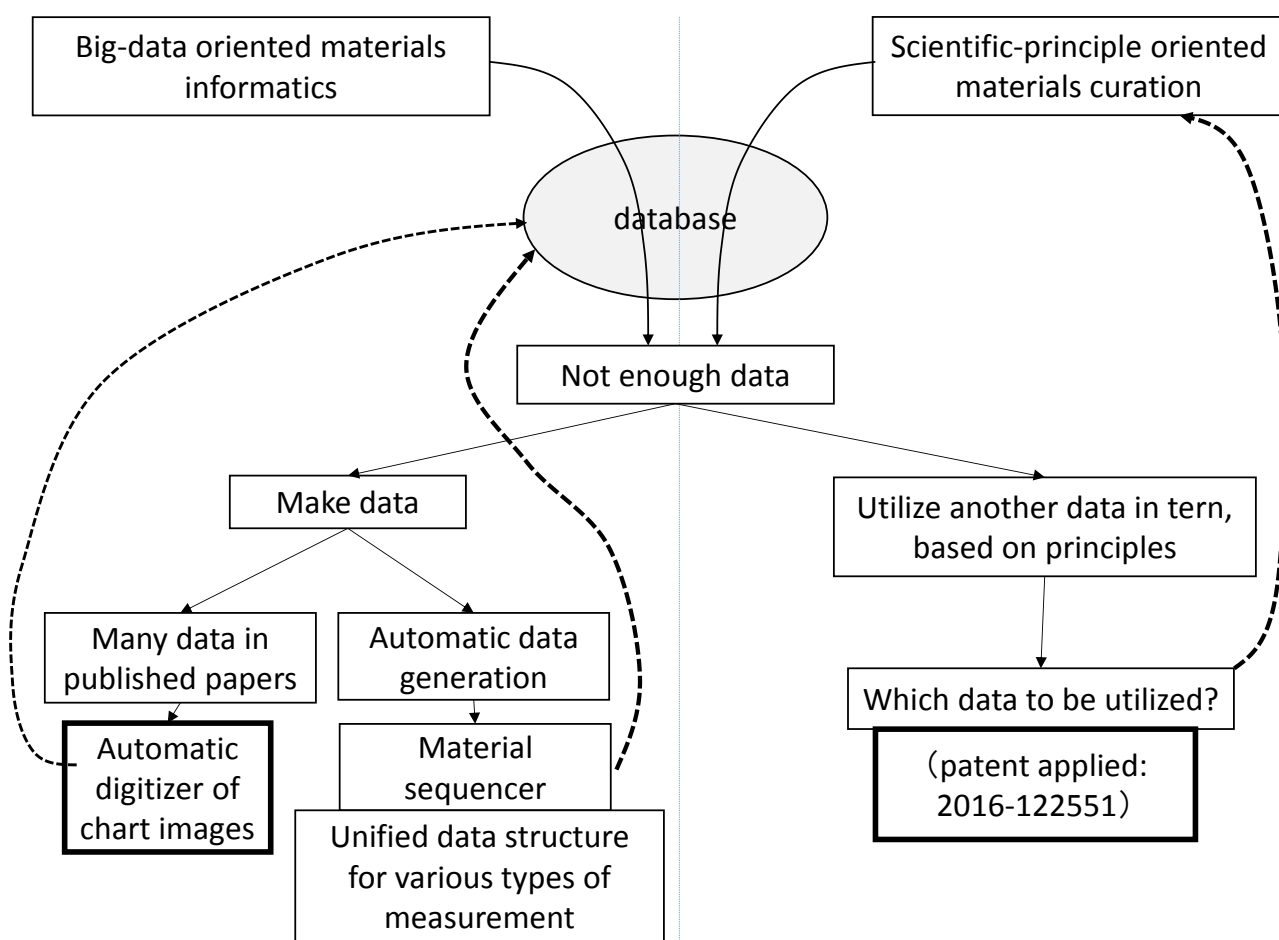
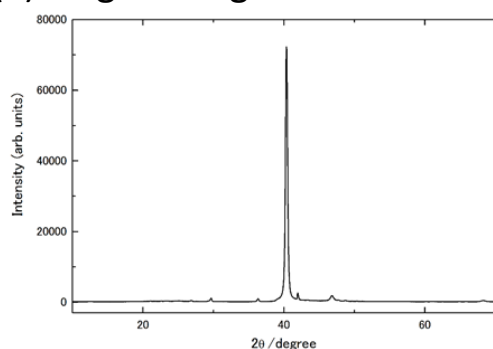
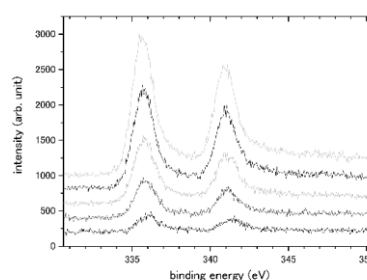


Fig. 1. Overall view of data-utilization project conducted by the author. The left side is concerning data generation, while the right side is concerning utilization of scientific principles,

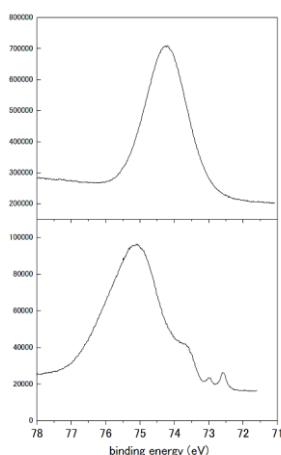
(a) Single straight line



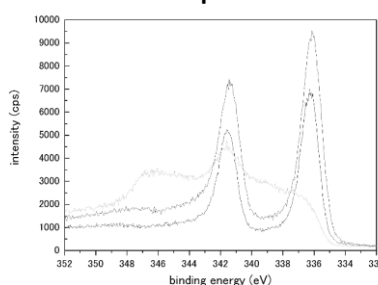
(c) multi-lines without overlap



(b) multi-column single lines



(e) multi-colored lines with overlap



(d) dotted-line & broken line

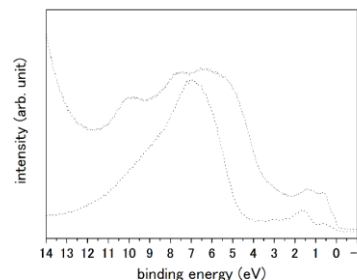


Fig. 2. Types of spectral image, which can be handled with the software.

部門を通じて有償供与されている（特許技術許諾に類似の形態）[2].

この数値化ソフトを用いて類似材料のスペクトルデータを数値化し、自身の測定スペクトルの数値処理（スペクトル差し引きなど）に用いたり、数値化したデータを機械学習の学習用データとして用いて自身のスペクトル同定に活用したりすることが可能である。

3. 科学原理の利用

3.1 背景

科学原理を知識データとして利用する方法は、著者が提唱してきた「マテリアルキュレーション™」＝「望む特性を持つ材料を、従来の延長線上ではない材料系で探索するための方法論」の一部である。「マテリアルキュレーション™」の基本は、「望む特性」が、本当にそれが必要なのかを、なぜその特性を望むのかに遡って検討し、本当に必要なもの＝wantsを見つけることからスタートする。本当に望む特性が規定されたなら、その特性が、科学原理的

に何によって決定されるかを調査し、望む特性を得るために原理的にどのような条件が必要かを絞り込む。絞り込まれた条件ごとに、その実現可能性を、データベース中の計測・計算データやインフォマティクス技術を利用して検証し、実現可能性の高い方法を見出す。このプロセスを模式的に示したのが Fig.3 である。

3.2 科学原理を利用した材料探索

Fig.3 の 2), 3) の過程が、望む特性が、科学原理的に何によって決定されるかを調査し、望む特性を得るために原理的にどのような条件が必要かを絞り込む部分である。この過程では、望む特性の決定要因を分解する過程で、他の特性との関係が明らかになる。その一例を Fig.4 に示した。これは、遷移金属炭化物において、炭素欠損（欠損はバルク・表面の両方で考慮）が仕事関数に与える影響（組成比が 1 : 1 から炭素不足の方へずれた時に仕事関数が増加するのか減少するのか）を予測する方法を見出す[3]のに用いられたものである。仕事関数は表面とバルク

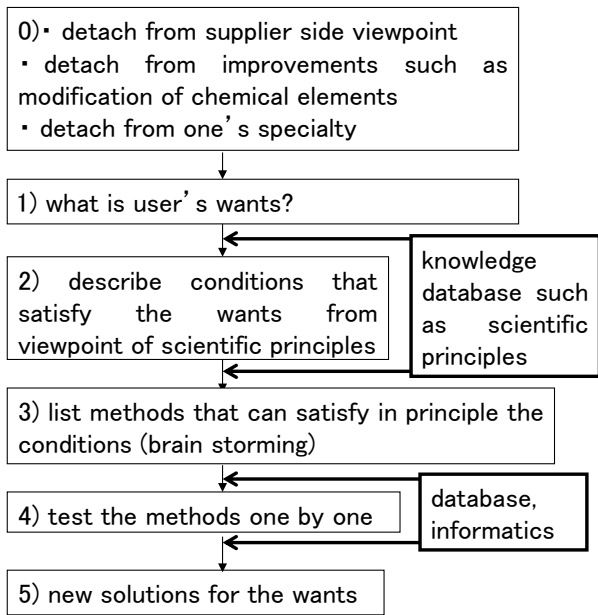


Fig. 3. Schematic representation of the process of "materials curation".

の両方の影響を受け、それぞれの影響が何によって決定されるかを科学原理に基づいてたどってある (欠陥をバルク・表面の両方で考慮するので、表面組成・バルク組成との関係が存在). 特に注目されるのは、バルク中の電子のエネルギー安定化は固体中の結合の強さに依存し、それが (遷移金属炭化物の系においては) ビッカース硬度という機械的な特性の決定要因でもある点である. この例では、材料系がハードコーティングに用いられる材料であったため、炭素欠損がビッカース硬度に与える影響についての実験データと考察が過去に豊富に存在する. ビッカース硬度が炭素欠損によりどう変化するかを予測する方法が存在しており、それを用いることで、仕事関数が炭素欠損によりどう影響を受けるかも予測することができる.

3.3 様々な特性の関係性

3.2 では「仕事関数」と「ビッカース硬度」の様に一見何の関係も無さそうな材料特性の間に、科学的原理に基づく関係性が存在することを利用した例を示した. このような関係性は、様々な特性の間に存在する. 様々な特性間の関係性をネットワーク図的に表現した例を Fig.5 に示した. Fig. 4 は、Fig. 5 のような関係性のネットワーク図のうち、仕事関数に影響を与える特性とビッカース硬度の部分のみ抜き出したものとみることができる.

著者は、このような関係性を利用して、これまで

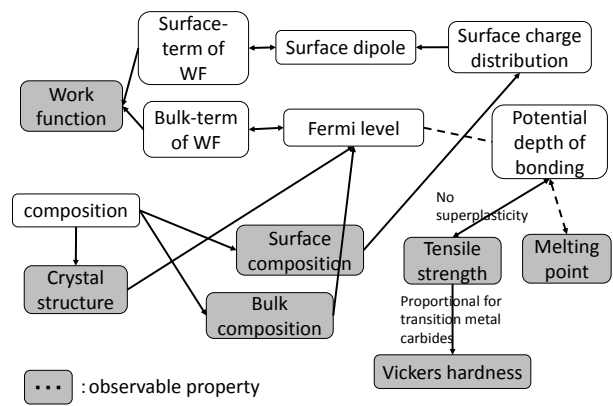


Fig. 4. Relations among various materials quantities that connect work function and hardness and the specific fields of science where the relations are well known.

に表面偏析の予測法を開発したり、金属-酸化物界面の界面結合の原子種を予測する方法を開発したりしてきた[4]. その他の様々な特性予測のためには、非常に多くの関係性が Fig. 5 の様にネットワーク化されることが必要である. また、非常に多くの特性が関係によりつながれば、人間が目で見つなかりを確認することはできない. そこで、このようなネットワーク構造を数学的に解析する既に確立した手法である「グラフ理論」[5]のアルゴリズムを用いることを考えた. グラフ理論を用いると、何と何がつながっているかを簡単に調べることができる.

3.4 関係性のデータベースと探索システム

3.3 で述べた様々な特性の間関係性を集めてデータベース化し、グラフ理論による関係性の探索ができるシステムのプロトタイプを構築している [6]. Fig. 6 にプロトタイプ概念図を示した. Fig. 5 では特性同士をつなぐ線に矢印を描いていないが、

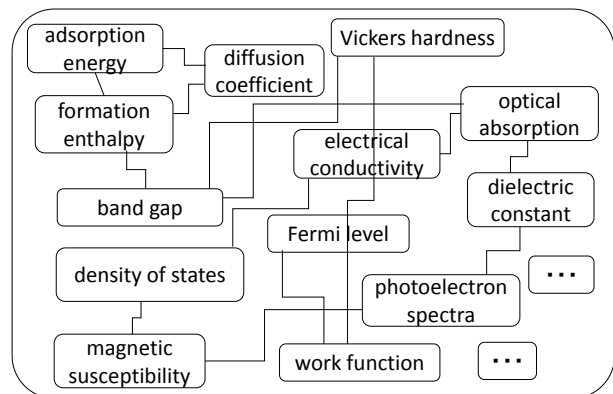


Fig. 5. Schematic relation network diagram among various physical properties.

多くの場合 Fig. 4 に示したように関係性には方向がある。両方向に関係性が成立する場合にはそれぞれの向きの矢印を描き、二つの特性の間に向きの異なる2本の線として表現する。このように表現すると、関係性は、「始点と終点（矢印の向きは始点→終点）が結ばれているかどうか」という情報を大量に集めたデータベースとして保存可能である。始点と終点の組み合わせに対し、関係性の有無を入力できるフォームを用意した。また、ここでは詳細は述べないが、関係性にも、厳密に理論的に定量的に成立するものから近似的に成立するもの、定性的に成立するものなど、様々な種類がある。入力フォームにはそのような関係性の種類と、定量的な関係式などが入力できるようになっている。また、科学原理が明らかで無くても、実験的に確立している相関は、関係性の種類にラベルを付けて入力できるようになっている。このような関係性のデータは、マクロによる重複などの入力チェックを得てデータベースの形で保存され、必要な時にグラフ表現へと変換できるようになっている。データベース上の関係性を探索する場合には、グラフ理論のアルゴリズムを用いて

行う。基本的に、①ある特性と別の特性との間に関係性があるかどうか、ある場合にはどのようなルートをとって関係しているか、②ある特性と関係する特性にはどのような特性があるか、の2種類の探索方法がある。また、関係性の探索には、上に述べた関係性の種類の違いによる重みづけや、報告されている実験・計算データの量による重みづけなどに応じた探索ができるようになっており[7]、材料探索の際に重視する観点によって適切と判断される材料が異なるようになっている。

4. 最後に

科学原理を利用した材料探索方法においても、探索方針を導き出した後に現実的な材料系を絞り込む際に実験・計算データは重要である。したがって、ハイスループット合成・評価やスーパーコンピュータを活用した自動第一原理計算などによる効率的なデータ生成、大量のデータの迅速な検索システムなど、実験・計算データとメタデータのデータ周辺の技術整備も非常に重要である。Fig. 7 には、科学原理の利用に立脚する「マテリアルキュレーション

<Relation data input side>

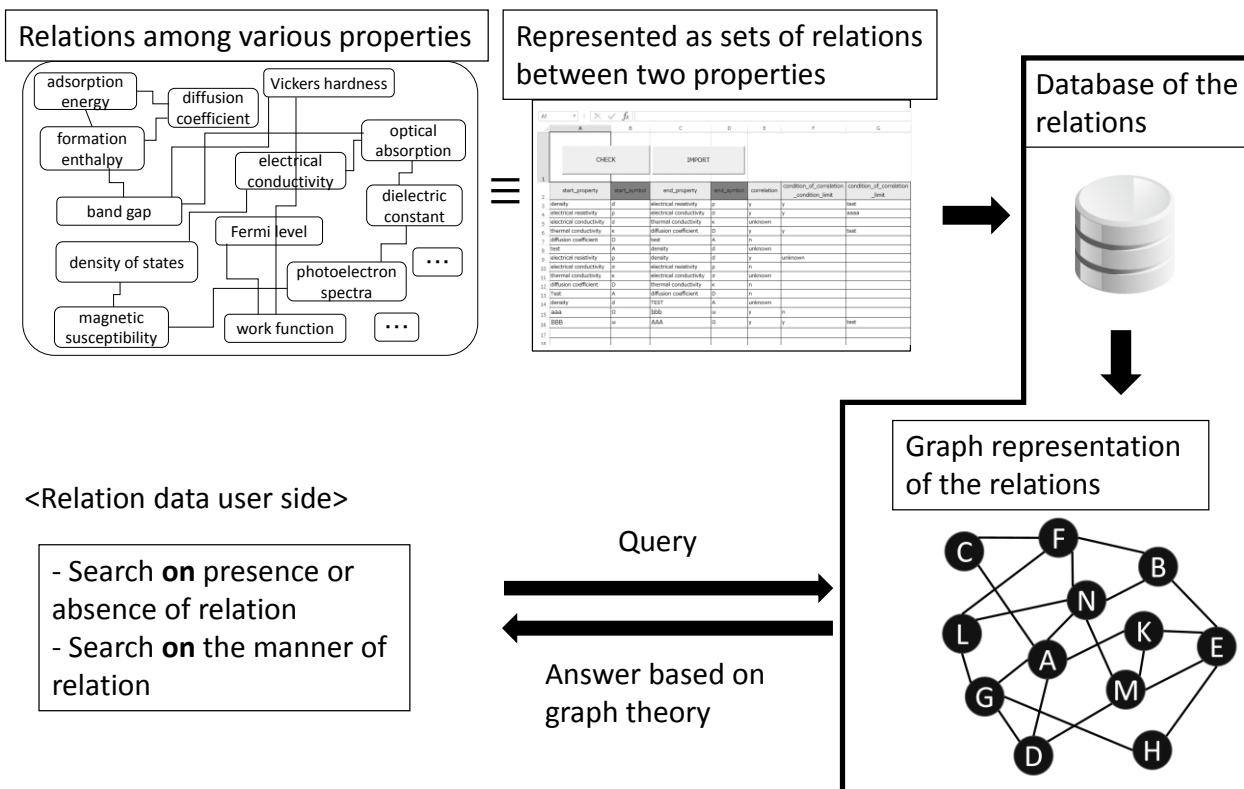


Fig. 6. Schematic illustration of the searching system of relations among various materials quantities.

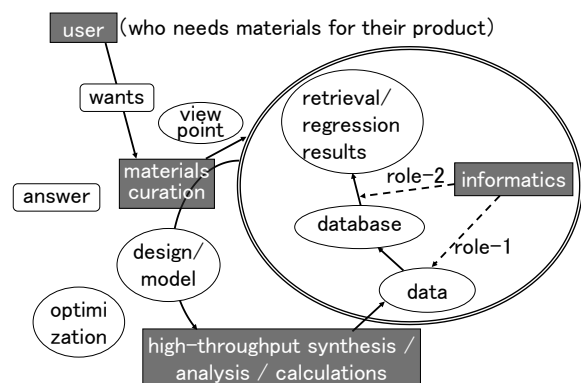


Fig. 7. Schematic of the relationship between database, informatics, materials curation, and high-throughput techniques for material development in future.

ン™」＝従来の延長線上にない材料の探索，から見た効率的データ生成やデータベースシステムとの関わりを模式的に示した．真の望みに基づいて必要な材料特性を抽出し，データベースのデータ（シミュレーションを含む）を活用して材料探索・設計指針を絞込み，ハイスループット手法を用いて実用的な最適化を図り，桁違いに効率的に材料探索を実現する，というのが将来の材料開発のあり方として著者が思い描く姿である．

5. 参考文献

- [1] 表面科学誌特集：マテリアルズ・インフォマティクス―表面科学のビッグ・データの構築一，*表面科学*, **36**, [10] (2015).
- [2] 物質・材料研究機構 外部連携部門 e-mail to: Technology-Transfer@nims.go.jp
- [3] M. Yoshitake, *J. Vac. Sci. Technol.* **A32**, 061403 (2014).
- [4] 吉武道子, *日本金属学会誌*, **80**, 603 (2016).
- [5] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, Elsevier, North-Holland (1976).
- [6] 吉武道子, 桑島功, 柳生進二郎, 知京豊裕, 特願 2016-122551.
- [7] 吉武道子, 桑島功, 柳生進二郎, 知京豊裕, 特願 2017-037387.

査読コメント，質疑応答

査読者 1. 鈴木峰晴（物質・材料研究機構）

交通網利用者情報やネット通販販売・購入情報に比べて，材料特性はまだしも，計測データを数理統計的に議論するには蓄積データ数が圧倒的に少ない問題を日々感じているところです．本報告では，画像化されたグラフやスペクトルを実際に利用できるように数値化する技術，科学原理にもとづいて統計処理を支援して材料探索を容易にする技術が簡潔に述べられています．計測（分析）を主に扱っている研究者・技術者が中心読者である J S A 誌に掲載されることは非常に役立つことだと思います．いくつか，読者の理解のために質問させていただきます．

[査読者 1-1]

画像自動数値化の際の分解能は，元々の図の点や線の太さ，読み取りの際の解像度で決まると考えてよろしいでしょうか．

[著者]

はい，元の画像の分解能で決まります．

[査読者 1-2]

科学原理を統計処理に利用する際には，関連する物性は「相関関係」でよろしいのでしょうか．それとも「因果関係」を有するものが必要でしょうか．

[著者]

科学原理の利用は，「原理」なので「因果関係」です．ただ，Fig. 5 のような関係図を作成していく上では，（詳細は次の解説記事に譲りますが）「相関関係」も，「因果関係」とは区別できる形で受け入れる構造になっています．

[査読者 1-3]

前項に関連しますが，2.1 で「その特性が，科学原理的に何によって決定されるかを調査し，望む特性を得るために原理的にどういう条件が必要かを絞り込む」，また「どう変化するかを予測する方法が存在して」と述べられています．特性を得るための条件が不明なときや変化を予測できないときには，統計処理に科学原理を繰り込むことは難しいということですか．それは，前項の質問の「相関関係」では難しいということになるのでしょうか．

[著者]

まず、科学原理に基づく手法では、統計処理はせずグラフ理論による解析を利用します。その点は誤解がないようにお願いします。

特性を得るための条件が不明なときや変化を予測できないときには、科学原理を繰り返すことで、別のどの特性の影響を受ける可能性があるかなどの知見は得ることができそうですが、その影響が、どの程度及ぶか、望ましい方向に影響が及ぶかなどの予測はできない、ということになります。

査読者 2. 匿名

[査読者 2-1]

Bulk composition がビッカース硬度や仕事関数に影響することは当然のことですが、炭素欠損は composition としては非常にわずかな変化ですので、材料によっては硬度（もしくは仕事関数）に与える影響は小さいのではないのでしょうか。炭素欠損の影響はゼロではないことは事実ですが、その影響が議論の対象になるレベルか誤差レベルか、更に炭素欠損がその影響に対する主要因であるか否かが重要で、読者もそのような疑問を持つのではないかと思います。具体的な材料での事例を挙げて、炭素の濃度の変化、それに伴う硬度や仕事関数の変化の数字を挙げた説明ができないでしょうか。

[著者]

お尋ねの点は、本稿の目的からは枝葉に当たり、論文として既に発表しており参考文献[3]として引用していますので、この関係性に特に興味を持つ読者には参考文献をお読みいただければと存じます。

[査読者 2-2]

上記にも関連しますが、プロパティや物性同士の影響度がゼロというケースはないと思いますが、非常に小さい場合と、顕著な場合があるかと思えます。Fig. 4 のネットワーク図は、その影響度を表すことで初めて実用的になるかと思えますが、この図では影響度の大きなものだけを掲載するという考え方でしょうか。

[著者]

Fig. 4 のネットワーク図はあくまで説明用のもので、より実用的なネットワーク図に関しては、「物性間関係図」として本稿執筆時点では投稿中だった原

著論文（本稿掲載決定時点で印刷中）に記述しております。関連内容の解説の執筆を JSA 誌から依頼されており（5 月末脱稿予定）、そちらに記述予定です。