

平成 21 年 5 月 29 日現在

研究種目：基盤研究(B)

研究期間：2005～2008

課題番号：17300030

研究課題名（和文） データ駆動型材料設計システムに関する基盤研究

研究課題名（英文） Basic Research on Data-driven Material Design

研究代表者

岩田 修一（IWATA SHUICHI）

東京大学・大学院新領域創成科学研究科・教授

研究者番号：50124665

研究成果の概要：データ駆動型の方法論の基礎を提示するため、データサイエンスの視点から検討した。材料データを記述するための変数群：メタデータは複雑で、精度や粒度も多様であるため、それらの非均質性に適応可能な測度論を検討し、データシステムの基本的な枠組みとした。データは、測定対象の属する空間あるいはメタデータが張る空間の部分集合についての「何か」をはかった結果についての記述であり、部分集合の測度とよばれる。データ駆動型材料設計は、データを集積することによって、社会のニーズに対応した解空間を作成し、ニーズに最も適合した材料の組成、諸構造、特性、価格を特定する設計作業と定義した。測度については、観測・測定方法あるいは経済性の限界に依り不完全であるため、多様なデータ群を目的に沿って誤差を補正し不完全な部分をモデルや近似により補完し、ニーズに合うデータ群を探索（写像）するプロセスを、材料データベースをプラットフォームにして実装し、データ駆動型材料設計の実例の蓄積を Web 上で展開した。

交付額

（金額単位：円）

	直接経費	間接経費	合計
2005 年度	3,600,000	0	3,600,000
2006 年度	3,200,000	0	3,200,000
2007 年度	3,200,000	960,000	4,160,000
2008 年度	3,200,000	960,000	4,160,000
総計	13,200,000	1,920,000	15,120,000

研究分野：総合領域

科研費の分科・細目：メディア情報学・データベース

キーワード：データ駆動型、材料設計、設計手法、データマイニング、反経験的手法、第一原理計算、ソフトウェアモデル、材料メタデータ

## 1. 研究開始当初の背景

科学技術超振興調整費：物質・材料設計のための仮想実験技術に関する研究で、第一原理的手法から連続体モデルまでを物質・材料や特性の違いに合わせて統合するための基盤技術を確立し、電子機能、構造機能についての設計事例をデータとモデルを統合したタスクフローとして定義し、誰もがインターネット

トによりアクセス可能な情報資源の形にした。

科学技術振興調整費：「人間系の特性を考慮した大規模・複雑システムのモデル化、解析、制御、設計に関する統合の研究」で、新しい人工物につながる設計作業を論理的な無駄のない定式化可能な手順と非論理的な発見的探索とに分離し、前者をデータベースに対する論理操作、後者を予測普遍性の高い手法の効

率的な実行による仮設形成と適合する解の選択操作の組合せで情報システムとして実現することを試みた。

文部科学省研究基盤研究(B)一般:「インターオペラビリティに関する研究」で、XMLを活用したデータ表現のシンタックスに関する検討及び言葉や論理が形成される初期過程に関する基礎的研究をベースにしたセマンテックスについての系統的な取り扱いを検討し、多数の研究者の参加によりネットワーク上での情報共有が効率的に進められるような条件を整えた。

文部科学省研究費補助金(国際学術研究):「材料設計のための国際的に分散したデータベースおよび知識ベースの統合化」により、材料データベース構築のための国際プロジェクトを組織し、2元系非有機系材料の基礎データを格納した材料設計のためのプロトタイプシステムを構築し、本研究のために基盤を構築した。

以上の準備を基に本研究を提案した。

## 2. 研究の目的

(1)設計の醍醐味は、通常の解とは質的に異なる予想外の属性と価値をもった解の獲得にある。本研究の目的は、近年、格段に充実しつつあるデータベース、モデリング手法、実験管理手法を徹底的に活用した新しいタイプの設計システムを材料を例に構築し、最適解のみならず予想外の結果を含む多様な設計解の創出が可能な設計システムの範例を提示する。

(2)材料に関する大規模なデータベースを基盤とした材料設計手法だけでなく、他分野にも適用可能な普遍性の高い求解手法の開発をする。

(3)これまで長年の研究連携の成果として準備した材料設計の基盤となる本格的なデータベースを出発点として、本研究では、この技術要素の改良、設計支援機能の拡充、ネットワーク化を実現する。

## 3. 研究の方法

物質・材料の設計に関するデータマイニングと知識発見(Data Mining and Knowledge Discovery)について、下記の視点から、過去、約40年間の代表的事例を以下の項目に関して編集、デジタル化する。

(1)構造・形態と特性・機能との間の相関。

(2)構造・形態、特性・機能についての測定あるいは記述の粒度の向上。

(3)構造に関する階層性の導入:同位対比、電子構造、結晶構造、ミクロ組織、他のメゾ構造、マクロ構造/高次構造、部品、構成要素、全システム、環境を視野に入れる。

(4)特性・機能は、多くの場合、時間依存、空間依存性を有するため、それぞれの時空間

のスライスで構造との相関を考える。

(5)計算可能なモデルは絶対零度、平衡状態等の理想状態に関するものが多いため、実用上重要な室温から融点、非平衡状態におけるミクロからマクロへのシームレスなマルチスケールシミュレーションの補完的な近似方法を開発する。

(6)設計問題を順問題解法だけでなく、逆問題解法を併用し、設計解への収束性を向上させる方策を考える。

(7)設計知識の発見は、単純に構造-活性相関のパターンを選択すれば良い場合から、未知のパターン探索、さらには新規データの獲得をとまなう設計戦略等定という高度なレベルを考える。

(8)問題が複雑で必要な情報が不十分な場合、データの獲得、評価、パターン発見、設計シナリオの作成を組み合わせ、いわゆる“戦略的”な問題解決を実施する。

(9)材料設計に良く活用された構造パターンに関する因子例は、 $e/a$ 、禁制帯巾、 $M_d$ 、電気陰性度、電子空孔密度、格子定数ミスマッチ、原子間ポテンシャルパラメータ、各種特性温度、配位数、最調密面、状態図パターン、テクスチャー、アスペクト比等々である。こうした基本的なパラメータによる構造-特性相関については、ライブラリとして準備する。

(10)材料開発の現場では、目標とする材料の開発に至る時間が重要となるため、ニューラルネットワーク、自己組織化マップやラフセットのような現象論的であっても簡便な相関導出手法を活用する。

(11)複雑システムの属性は時間の経緯とともに変動することが多いため、ロバスタな制御、モデリングが要請される。近年、ドラックデザインの分野で着目されているコンビナトリアル化学、コンビナトリアル合成がそうしたスピード感のあるアプローチの代表例であり、そうしたダイナミックなアプローチの適用を検討する。

## 4. 研究成果

(1)回帰分析やクラスター分析などの一般的な多変量解析手法により、線形モデルによる第一近似としての現象論的な構造-特性相関を抽出するためのツールを開発した。

多次元の変数を持つデータ群から、ある種の規則性を抽出しようとする場合、適当な座標軸を設定し、より低次元の空間にデータを射影し、そこでの相関を解析するのが通常の方法である。しかしながら一般的には低次元化する際の座標軸の設定に不明な部分が多く、高い相関係数が得られる座標を見つけ出すのは容易なことではない。そこで、ここでは標準的な部品展開が示されている系を選択し、部品間の相関を求めることを考える。考え方

としては最初から複雑な系の解析を試みるのではなく、高品質のデータを準備しやすく、データが揃っている基礎物性、構造要素の属性についての整理を行い、逆問題アプローチで使用する試行関数を準備し、段階的な問題解法を試みた。事前の検討では、物質・材料の約60の基本属性から5パターンの相関が帰納的に導出されていたが、この事例を参照して相関抽出のための基本アルゴリズムを汎用ツール化した。

(2) 既往の論理式、経験式について、データベースから導出することを試み、物理的知見から合理的と考えられる相関関係の記述、データベースによる帰納とがほぼ等価であることを確認した。重要な点は、データの品質管理にある。

(3) 構造・活性相関を活用するプラットフォームとして規則性探索のための構造マップを構築した。

実際にここで用いるマップの軸としては Mendeleev Number (MN) を基本独立変数とし、上記の相関関係を活用しながら試行関数を準備し、化合物構成元素を軸とした空間上での結晶構造の分布、AET (Atomic Environment Type: 局所的な原子配置環境、つまり結晶中のある原子を中心とし、その周りの近接原子の配位構造に対応する分類方法) マップ表現による規則性の発見的探索が有効であることを確認した。

(4) 以上の具体例の検証を通してデータ駆動型材料設計システムの基本仕様を決定した。

さらに当該分野の新たな知見を反映するためのデータ記述、モデル記述、可視化・編集のためのインターフェイスの仕様を決定した。

(5) 相関関係の探索のために開発したプラットフォーム上に既往の微視的～巨視的モデルを実装し、データ、モデル、経験則を組み合わせた順問題解法の高速化、収束性の向上を計った。

(6) 構造マップの本質はデータの高品質化と非明示的な解の創発にあり、構造マップ上での多様な解の比較・対比・編集、分類・汎化・集約によるアドホックな経験則の獲得、普遍性のある原理による近似を通して、複雑なシステム間の類推の可否についての検討を行った。静的な構造に関する比較は可能であるが、実用上重要な格子欠陥のふるまいや、相変化等の特異点近傍のふるまいに関しては、想定される複雑なふるまいを解明するための課題の大きさを再確認しただけで、挑戦の糸口が明らかになったに過ぎない。

(7) 構成要素と全体との間に存在する非明示的な関係を求める逆問題解法をモデルの選択/作成とパラメータフィッティングに分けてインタラクティブに設計解を探索するためのインターフェイスを開発した。

(8) 構造マップやデータ評価のためのフィ

ルター群を拡充して、ネットワークを介して収集した多種多様な大規模不完全情報を俯瞰し、品質管理し、再活用するための手法の検討開と、上記の類推を活用した効率的な補完手法についての検討を行った。

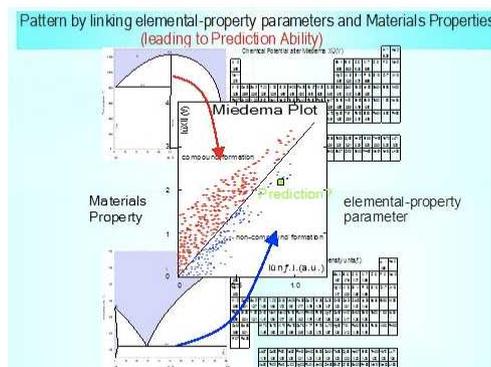
(9) 環境調和型的人工物を例題として選び、求解に至るデータ、モデルの品質管理手法、特に大規模データベースの俯瞰・可視化と帰納による“常識”の提示と、相関関係抽出、事実の多面的な解釈を創発するインターフェイスを設計、開発し、利用可能な装置群、情報システム群をプラグイン型で統合化するためのメタデータ、人間系の試行錯誤を管理、支援するためのインタラクティブな知的環境を構築した。

(10) 複雑システムの属性は時間の経緯とともに変動することが多いため、ロバスタな制御、モデリングが要請され、ダイナミックなアプローチが必要となるが、誤差の蓄積に起因して問題解決が容易でないとされている時間発展型の課題についての解法を検討した。ドックデザイン分野で着目されているコンビナトリアル化学、コンビナトリアル合成等のスピード感のあるアプローチを開発している米国のグループとの研究協力を実施し、ネットワーク状で共同研究を実施し、開発した材料設計プラットフォームの可能性と限界を明らかにした。

以上の4年間の成果をシステムとして公開し、俯瞰と共創の可能な構造マップ型のインターフェイスの構築により、ユーザー参加型の自立的な情報プラットフォーム構築のための基礎資料とした。

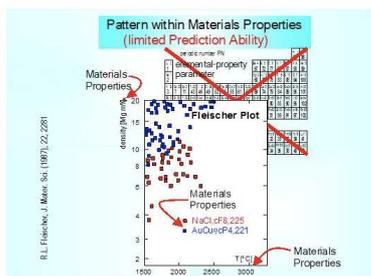
公開したプラットフォームのスナップショットを以下に示す。

prediction for chemical systems where no experimental information is known, a chemical potential difference, as well as its size difference, and therefore its lo compound forming area are like also to form new compounds.



Key targets in materiel design

the fundamental physical materials properties, or to its elemental property parameters. Fleischer and Mieder successful to discover 'strong patterns' within large groups of materials data.



Fleischer's graph plots the density versus the melting point of some hundred binary intermetallic (intermediate having very high melting points) compounds. From his figure one can easily distinguish two regions the upper represent some binary compound having the same crystal structure (AuCu prototype) and all the red dots are (NaCl prototype). Obviously, this shows a pattern which separates two kinds of groups likely to have similar request to plot materials properties of each compounds, one needs to know from experiments, e.g. in Fleischer's point of each compound. In other words a 'new' material has first to be found, then its physical properties have

このようなプラットフォームの共有は背景知識の学習を必要としない関連分野の専門家間では極めて有効であることを確認し、そうした経験を通して将来の”Data-centric Science”へのパースペクティブを得た。

## 5. 主な発表論文等

〔雑誌論文〕(計10件)

岩田修一、CODATAの活動と展望について：Data-centric Science and Societyへ、学術の動向、5月号、79-84、2009. 査読なし

N. Kiselyova, S. Iwata, V. Dudarev, I. Prokoshev, V. Khorbenko, and V. Zemskov, Integration Principles of Russian and Japanese Databases, International Journal "Information Technologies and Knowledge", Vol. 2, 2008, pp366-372, 査読有り

J. Jia, S. Iwata, Y. Seida, Y. Kaneta, N. Oka and Z. Zhang, Fractal Characteristics of Colloid Deposition, Data Science Journal, Vol. 6, 2007, pp.S206-S219 査読有り

H. Wang, Y. Chen, Y. Kaneta and S. Iwata, First principles investigation of the structural, electronic and optical properties of olivine-Si<sub>3</sub>N<sub>4</sub> and olivine-Ge<sub>3</sub>N<sub>4</sub>, Journal of Physics: Condensed Matter, Vol. 18, pp10663-10676, 2006. 査読有り

T. Mohri, M. Ohno and Y. Chen, First principles calculation of ordering phase transition, J. Phase Equilibria and Diffusion, Vol. 27, 47-53, 2006. 査読有り

岩田修一、科学技術データとオープンアクセス、学術の動向、12月号、40-43、2005. 査読なし

岩田修一, Science and the Digital Divide, SCIENCE, Vol. 310, pp405, 2005. 査読有り

Ying Chen, First-principles investigation of phase equilibria of Fe-based alloy systems, TMS, pp663-668, 2005. 査読有り

T. Mohri, Y. Chen and M. Ohno, First principles calculation of ordering phase transition, TMS, pp633-650, 2005. 査読有り

M. Ohno, Y. Chen and T. Mohri, Hybridization of phase field and cluster variation methods towards first-principles calculation of microstructural process, J. Jpn. Soc. Mechanical Eng., Vol. 108, pp798-801, 2005. 査読有り

〔学会発表〕(計6件)

S. Iwata, Standard Reference Data of Tensile Properties for Heart-resistant/Structural Steels, CODATA Activities on Data Quality, 2008年10月29日, Daejeon

岩田修一, Overview on Materials Design, CODATA: "Kyiv School", Data Science and Design Science on Materials, 2008年10月, キエフ

S. Iwata, Institutional Design of Global Information, Commons for Asia The 7<sup>th</sup> Science Council of Asia Conference, 2007 June 14-16, Okinawa

S. Iwata, Genesis of E-Science and next step for its implementation, Consultation meeting on WSIS Action Line C7

'E-science', 2006年10月22日, 北京

S. Iwata, E-Science in GICSI Missing Links -, Japanese - German I&D Panel, e-Science Workshop, 2006 May 25-26, The University of Tokyo

S. Iwata, Global Information Commons  
for Science Initiative, World Summit on  
the Information Society,  
2005年11月16-18日, チュニス

〔図書〕(計1件)

S. Iwata, Y.Ohsawa, S.Tsumoto, N.Zhong,  
Y.Shi and L.Magnani, Springer,  
Communications and Discoveries from  
Multidisciplinary Data, 27 Aug 2008

〔その他〕

<http://ilab.k.u-tokyo.ac.jp/tiki/>  
(ID: guEST, Password: ena4eVa4)

## 6. 研究組織

### (1) 研究代表者

岩田 修一 (IWATA SHUICHI)  
東京大学・大学院新領域創成科学研究科・教授  
研究者番号: 50124665

### (2) 研究分担者

陳 迎 (CHEN YING)  
東京大学・大学院工学系研究科・准教授  
研究者番号: 40372403

金田 保則 (KANETA YASUNORI)  
東京大学・大学院工学系研究科・助教  
研究者番号: 00262048

### (3) 連携研究者