

令和元年6月21日現在

機関番号：33302

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K13739

研究課題名(和文) 進化的アルゴリズムとシミュレーションを活用した新規分子ワイヤ物質探索

研究課題名(英文) Research for new molecular wire utilizing genetic algorithm and numerical simulation

研究代表者

林 亮子 (Hayashi, Ryoko)

金沢工業大学・工学部・准教授

研究者番号：30303332

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：データマイニング適用の例題として、炭素、酸素および水素の3種類の元素のみを含む非常に基礎的な化合物群について発火点予測及び発火点決定ルール抽出を試みた。264件の学習データによるランダムフォレストを用いてテストデータ30件中26件で100K以内の誤差で発火点を予測することができた。決定木では30件中22件で100K以内の誤差で発火点を予測することができた。決定木では、これまで経験的に知られていた発火点に関するルールのいくつかを確認でき、さらにそれらのルールにおける閾値を定量的に調べることができた。今後は2原子分子のシミュレーション結果を用いて分子ワイヤにつながる材料設計を試みる予定である。

研究成果の学術的意義や社会的意義

近年材料科学において、これまで蓄積したデータとデータ科学を活用して新規の材料を開発する機運が国内外で高まっている。本研究はデータ科学とシミュレーションを組み合わせることで材料開発を目指す材料設計プラットフォームシステムを目的とし、そのシステムを用いて分子ワイヤの探索問題を解くことを目指し、必要となる研究開発を行うものである。分子ワイヤは分子レベルで作る結線であり、近年分子機械や分子回路が現実的に作成されるようになってきたために注目されている物質である。そのため、本研究が目的とする分子ワイヤ物質の探索は、ナノテクノロジーや材料科学の観点からも意義がある。

研究成果の概要(英文)：I studied rule extraction and prediction for the ignition point of simple chemical compounds including carbon, oxygen and hydrogen only using data mining technique such as decision tree and random forest, as a trial of data mining. I used fundamental material values and the number of characteristic structures as descriptors. The learning data includes 264 kinds of chemical compounds and the test data includes other 30 kinds. As the result, random forest predicted ignition point for 26 molecules within 30 molecules of test data within 100K of RMSE. Decision tree also predicted ignition point for 22 molecules within 30 test datas within 100K of RMSE. Moreover, decision tree also got several rules for ignition point mainly known empirically, and examined threshold on the rules quantitatively. Future works will be about material design using dimer simulation result.

研究分野：高性能計算, データマイニング

キーワード：計算科学 データ科学 データマイニング シミュレーション 発火点 分子ワイヤ 材料設計

### 1. 研究開始当初の背景

近年材料科学において、これまで蓄積したデータとデータ科学を活用して新規の材料を開発する機運が国内外で高まっている。本研究はデータ科学とシミュレーションを組み合わせることで材料開発を目指すものである。研究代表者はシミュレーションの高速化を専門としており、計算科学分野の研究者とも共同研究の経験があるが、その過程で結果データ処理の重要性に気づき、近年ではシミュレーションとデータマイニングを組み合わせさせた材料設計プラットフォームシステムに関する研究開発を行っている。本研究ではそのシステムを分子ワイヤの探索問題に特化し、必要となる研究開発を行うものである。分子ワイヤは分子レベルで作る結線であり、近年分子機械や分子回路が現実には作成されるようになってきたために注目されている物質である。そのため、本研究が目的とする分子ワイヤ物質の探索は、ナノテクノロジーや材料科学の観点からも意義がある。

### 2. 研究の目的

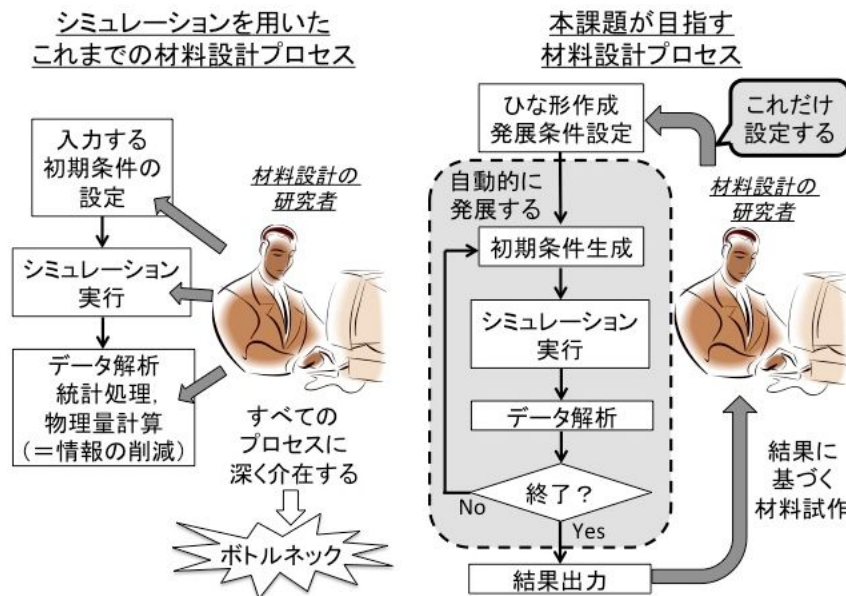


図1. 材料設計システムの概要

本研究が目指す材料設計プロセスの概要を図1に示す。これまでの計算材料科学は、シミュレーションの初期条件設定から実行、結果データ処理までの多くの場面で研究者の介入を必要としていた。しかし、人間の能力には限りがあるため、人力処理がボトルネックになる。一方、シミュレーション実行に関連する多くの作業は、近年では自動的に行えるため、研究代表者はシミュレーションによる探索プロセスの自動化を目指している。本課題では、分子ワイヤ物質探索を材料設計の実問題として扱う。

### 3. 研究の方法

**問題設定:** 本研究は、特に分子間力によって結合する分子ワイヤ物質の探索を目的とする。そのような物質は、化学分野で用いられる構造ジェネレータでは扱えず、分子を配置して構造を最適化しながら探索する必要がある。本研究では分子ワイヤを1次元的な分子の並びと見なし、進化的アルゴリズムを用いて分子パーツを並べた分子列を複数個生成する。それらの分子列に対して構造最適化を行い、安定な構造を作る分子列でかつ物理長が長くなり、電気伝導度が高い分子列を探索する。分子には向きと大きさがあるので、同じ分子列でも、初期条件における分子の配置は複数通りが必要である。本研究では乱数を利用して配置を生成するため、構造最適化の結果、分子の構造が初期条件から変化し、極端な場合は分子が壊れる可能性がある。

**扱う分子の条件:** これまで研究代表者が調査した結果、非常に精度の良い計算手法を用いて炭素原子10個程度の分子がデスクトップ計算機上で数分程度の実行時間で構造最適化できる。そのため、当初は本研究で扱う分子サイズを炭素原子10個程度以内に限定して研究を行う。分子ワイヤ物質には炭素原子5個程度の分子もあるため、このように制限しても大きな問題はない。

**本課題の開発範囲:** 本課題ではシミュレーションプログラム本体の開発ではなく、その入出力の扱いに力点を置くため、シミュレーションには既存のパッケージ、具体的には Gaussian09 を利用する。入出力を扱う手法は特定のパッケージに依存せず、入出力ファイルの形式に対応すれば他のパッケージにも拡張可能であるため、現段階でパッケージを特定しても汎用性は失わない。

#### 4. 研究成果

1. データマイニングの材料関連分野への応用可能性を検討するため、例題として発火点予測及び発火点決定ルール抽出を試みた。炭素、酸素および水素の3種類の元素のみを含む非常に基礎的な化合物群について一通りの調査を行なった。記述子は、分子の基礎的な量と炭素原子個数、酸素原子個数、さらに炭素、酸素、水素からできている化学物質の特徴的な部分構造(いわゆる「官能基」)を使用し、あわせて18個とした。データは全部で294件用意できたため、テストデータは10%程度の30件、学習データを残りの264件とした。そして、分類や予測を行う代表的なデータマイニング手法として知られている決定木とランダムフォレストを使用した。その結果、ランダムフォレストを用いるとテストデータ30件中26件は100K以内の誤差で発火点を予測することができた。決定木では、30件中22件で100K以内の誤差で発火点を予測することができた。また、決定木では、これまで経験的に知られていた発火点に関するルールのいくつかを確認でき、さらにそれらのルールにおける閾値を定量的に調べることができた。以上の内容を国内外学会における口頭発表、さらに情報処理学会論文誌「数理モデル化と応用」論文として発表することができた。この内容については、さらに扱う元素を増やして検討を続ける予定である。
2. 次の研究対象として、分子間力を扱うような内容で、材料関連分野の研究者、特に研究協力が興味を持つような、分子ワイヤに発展させる中間問題となる例題を調査した。そして、2種類の原子が1個ずつ結合した分子である「ダイマー」の結合距離を記述子に用いてシート状物質の物性を調べる問題に着手した。周期表上で83番まで程度の、非放射性元素で材料によく用いられるような種々の原子の組み合わせについて総当たりにダイマーの物性値を計算するスクリプトを開発した。今後はこの内容をさらに発展させていく予定である。

#### 5. 主な発表論文等

##### [雑誌論文](計1件)

林 亮子, 決定木を用いた基礎的有機化合物の発火点決定ルール抽出, 情報処理学会論文誌 数理モデル化と応用, 査読有, Vol. 10 No.3 pp.20-31, (Dec. 2017).  
ISSN: 1882-7780

##### [学会発表](計9件)

林 亮子, 中田 侑江, 発火点決定ルールを調査する決定木作成の実行時間, 情報処理学会研究報告ハイパフォーマンスコンピューティング(HPC), Vol.2016-HPC-154(12), 2016, pp.1-8.  
ISSN: 2188-8841

林 亮子, 中田 侑江, 発火点決定ルールにおける決定木を用いた重要な記述子抽出の試み, 第39回ケモインフォマティクス討論会 講演要旨集, P7, (2016).

林 亮子, 決定木を用いた基礎的有機化合物の発火点予測, 情報処理学会研究報告 数理モデル化と問題解決(MPS), Vol.2017-MPS-113(9), 2017, pp. 1-6.  
ISSN: 2188-8833

Ryoko Hayashi, A Decision Tree of Ignition Point for Simple Inflammable Chemical Compounds Data Mining and Big Data, Second International Conference, DMBD 2017 Proceedings, Lecture Notes in Computer Science, vol 10387, pp. 165-172, (2017).  
ISSN: 0302-9743

林 亮子, 分子シミュレーションへのデータマイニング技術応用, 第31回分子シミュレーション討論会 講演要旨集, 105P, (2017).

林 亮子, 決定木とランダムフォレストを用いた発火点予測および発火点決定ルール抽出, 第40回ケモインフォマティクス討論会 講演要旨集, P05, (2017).

Ryoko Hayashi, Rule Extraction and Prediction for Ignition Point Using Decision Tree and Random Forest, ICPAC 2018, (2018).

林 亮子, データ科学と計算科学の協働による分子設計を目指して, 第41回ケモインフォマティクス討論会講演要旨集, 2C07, (2018).

林 亮子, 分子シミュレーションとデータマイニングの協働の試み, 第32回分子シミュレーション討論会講演要旨集, 216P, (2018).

#### 6. 研究組織

##### (1)研究分担者

(なし)

##### (2)研究協力者

研究協力者氏名: 水関 博志

ローマ字氏名：(MIZUSEKI, hiroshi)

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。