



2018年5月8日

報道関係者各位

慶應義塾大学

## ディープラーニングによる分子シミュレーションデータの高効率化 ー短時間のデータだけで、長時間のふるまいがわかる夢のAIー

慶應義塾大学大学院理工学研究科の遠藤克浩（修士課程1年）、友部勝文（2018年3月博士課程修了）、および理工学部の泰岡顕治教授の研究グループは、ディープラーニングにより短時間の分子シミュレーションデータを学習することで、長時間の分子シミュレーションデータを生成可能な新規モデルを提案し、いくつかの検証実験を通してその有用性を示しました。

分子シミュレーションは、分子の動きを再現することができる手法であり、生体物質、高分子、材料等、その利用先は極めて多岐にわたり、新規材料開発や病理解明に用いられています。一方で、大きな分子や長時間のシミュレーションを行う場合は、大規模な計算リソースを必要とするため、計算が困難という欠点がありました。本研究では、人工知能（AI）（※1）の一つであるディープラーニングにより、短時間のシミュレーションデータを用意するだけで、長時間のシミュレーションデータを予測することができる新規モデルを提案しました。これによって、分子シミュレーションを用いて研究を行っている企業や研究機関は、計算しなければならないシミュレーション量を減少させることができるため、研究開発の大幅な効率化を行うことが可能となります。さらに、本提案は分子シミュレーションデータだけでなく一般的な時系列データに応用することが可能であり、自然言語処理、経済データ、モーションデータ等、様々な時系列データへの応用が期待されます。

本研究成果は2018年4月26日（現地時間）に第32回アメリカ人工知能学会（AAAI-18）のサイトにて公開されました。

### 1. 本研究のポイント

- ・分子シミュレーションは、分子の動きを再現できる手法であり、大変有用である一方、計算コストが高いため、大規模・長時間のシミュレーションを行うことが困難であった。
- ・本研究グループは、短時間のシミュレーションデータを学習させることで、長時間のシミュレーションデータを予測できる新規モデルを提案した。
- ・検証を行った高分子の分子シミュレーションにおいては、約20倍の効率化に成功した。

### 2. 研究背景

分子シミュレーションは、分子を形成する原子一つ一つの動きを再現することができる手法であり、現象の分子メカニズムを明らかにすることができます。そのため、分子シミュレーションが用いられている応用範囲は生体物質、高分子、材料等、極めて多岐にわたり、新規材料開発や病理解明を目指した研究に日々用いられています。しかし、計算システムの大規模化や長時間化に伴う計算コストの増大が分野全体のボトルネックとなっており、近年の問題として挙げられています。

本研究は、機械学習の一つであるディープラーニングによって、分子シミュレーションの効率化を行いました。

### 3. 研究内容・成果

#### ◆ 確率的時間発展と蓄積するバイアスの低減

本研究は、分子シミュレーションを確率論的時間発展（※2）としてモデル化し、確率論的時間発展のためのディープラーニングの新規モデルを提案しました。確率論的時間発展の問題点として、繰り返し時間発展を行うと、誤差が蓄積されてしまうことが知られています（図1（A））。提案モデルは、この蓄積される誤差を低減させる仕組みがあるため、繰り返し時間発展を行うことを可能としています。図1（B）は、ノイズを含んだ調和振動子の動きを学習させ、同じ動きを提案モデルで生成した結果を表していますが、既存モデル（図1（B）下）は提案モデル（図1（B）中）と比較して、誤差が蓄積している様子が見て取れます。

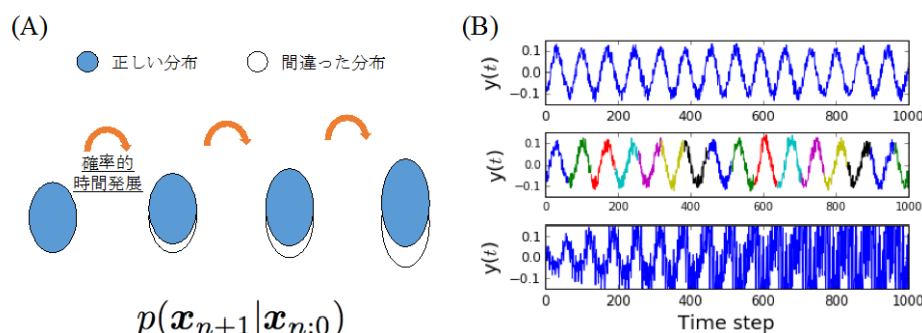


図1：確率的時間発展を繰り返すことで蓄積する誤差。(A) 蓄積する誤差の概念図。(B) ノイズを含んだ調和振動子を用いた実験。(B)の上から、教師データ（※3）、提案モデルによるデータ、既存モデルによるデータ。提案モデルにおいては、それぞれの色が一度に生成されたデータを表しています。提案モデルによるデータは、蓄積する誤差を低減できており、教師データとほぼ一致していることがわかります。

#### ◆ 高分子シミュレーションへの応用

高分子は、同じ分子が多数結合して構成されている大きな分子であり、車のタイヤやプラスチックに使用されている重要な物質です。高分子の物性を特徴付ける現象の一つに絡み合いがあります。これは、図2（A）のように高分子同士が絡み合う現象であり、高分子の拡散性や弾性を特徴付けることが知られています。この絡み合いが解ける現象は、短時間の分子シミュレーションでは再現不可能でした。本実験では、高分子の一つであるポリエチレンの分子シミュレーションを行い、短時間（640 ps）の分子シミュレーションデータから長時間データ（>500 ns）の予測を行いました。短時間の分子シミュレーションでは、絡み合いが解ける現象が再現されておらず、拡散性が通常より遅いことがわかります（ $\sim t^{0.5}$ ）。一方、図2（B）の提案モデルで予測された長時間データを見ると、絡み合いが解ける現象がしっかりと再現され、通常の拡散性になっていることがわかります（ $\sim t$ ）。

### 4. 今後の展開

本研究成果は、多くの分子シミュレーションを用いた研究に対して、短時間のシミュレーションデータを用意するだけで長時間のデータの再現を可能にするため、研究開発の大幅な効率化を促進することができます。また、提案モデルは様々な時系列データ（※4）に応用が可能であるため、

自然言語処理、経済データ、モーションデータ等、様々な時系列データの予測への応用が期待されます。

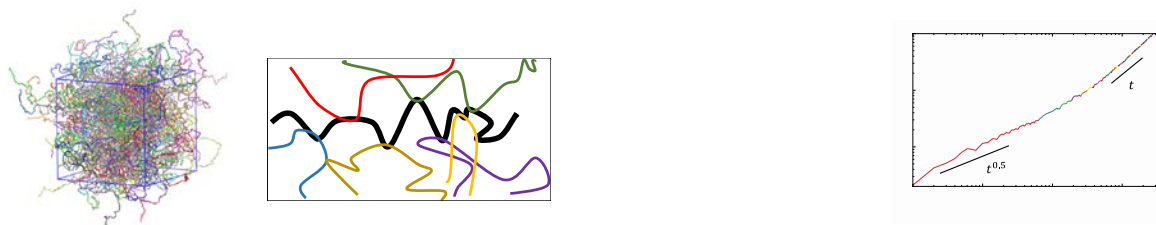


図2：高分子シミュレーションを用いた実験。(A) 実験で使用したポリエチレン系と絡み合いの概念図。(B) 左が、分子シミュレーションデータによるMSD(※5)、右が提案モデルによるMSD。提案モデルの図中には、実際に教師データとして使用したシミュレーションの長さをオレンジ色のバーで示しています。両対数グラフなので、時間が増加するほどバーは短くなります。

#### <原論文情報>

論文タイトル: Multi-step time series generator for molecular dynamics

著者: Katsuhiko Endo, Katsufumi Tomobe, Kenji Yasuoka

URL: <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16477>

#### <用語説明>

※1 人工知能(AI) : AIは、人間の知的活動を機械で再現する技術の総称であり、その中の一分野である機械学習は、データを解析し、有用な情報を抽出、活用する分野であると言えます。

※2 確率論的時間発展 : 分子を確率的に時間発展させることを指します。これに対して、本研究が対象としている分子シミュレーションの一つである分子動力学法は決定論的時間発展を行います。

※3 教師データ : AIを学習させるために準備された訓練用のデータ。

※4 時系列データ : 時間によって変化する連続的な値のデータのことを指します。代表的な時系列データとして、株価や心電図等が挙げられます。

※5 MSD : 平均自乗変位であり、運動の大きさ、つまり、粒子の拡散性を表します。

※ご取材の際には、事前に下記までご一報くださいますようお願い申し上げます。

※本リリースは文部科学記者会、科学記者会、各社科学部等に送信させていただいております。

---

#### ・研究内容についてのお問い合わせ先

慶應義塾大学 理工学部 機械工学科 教授 泰岡 顕治 (やすおか けんじ)

TEL : 045-566-1523 FAX : 045-566-1495 E-mail : [yasuoka@mech.keio.ac.jp](mailto:yasuoka@mech.keio.ac.jp)

#### ・本リリースの配信元

慶應義塾広報室 (竹内)

TEL : 03-5427-1541 FAX : 03-5441-7640

Email : [m-pr@adst.keio.ac.jp](mailto:m-pr@adst.keio.ac.jp) <https://www.keio.ac.jp/>