

jh250078

グラフニューラルネットワークと生成モデルを用いた 非晶質系動力学予測システム開発

芝 隼人（兵庫県立大学大学院情報科学研究科）

概要

研究では、乱雑な原子・分子配置を維持したまま固化した物質であるガラスの複雑な動力学解明に向け、機械学習手法の適用を推進した。具体的には、初期構造から動力学傾向を予測するために設計されたグラフニューラルネットワーク（GNN）「BOTAN」の改良と、時間発展シミュレーションへの生成モデルの応用という2面から研究を実施した。今年度は特に、温度変化に対応した改良モデル「T-BOTAN」を開発し、動力学の詳細な挙動を忠実に予測できるだけでなく、異なる温度への内挿的な適用も可能であることを実証し、原著論文の投稿を行った。また生成モデルに関しては、連続正規化フローを実装することで、物理的に適正な粒子配置の生成ができる段階までを検証・達成した。これらは、ガラスシミュレーションの将来的な改良や高度化に大きく寄与する。

1 共同研究に関する情報

1.1 共同研究を実施した拠点名

- 東京大学 情報基盤センター
- mdx I

1.2 課題分野

- 大規模計算科学課題分野

1.3 参加研究者の役割分担

- 川崎猛史（名古屋大学 ⇒ 大阪大学、副代表者）シミュレーション・物理モデリング
- 下川辺隆史（東京大学、副代表者）データ駆動モデル開発・アクセラレータ利用検討
- 華井雅俊（東京大学）データ駆動モデル開発
- 別所秀将（名古屋大学）シミュレーション・物理モデリング

- 朝比祐一（CEA Saclay）データ駆動モデル開発

2 研究の目的と意義

ガラスは、乱雑な原子分子（粒子）の配置を保ったまま固まった状態の物質である。広範な種類の物質が、液体状態からの急冷によって、あたかも相転移のように急激に固まってガラスとなる。ガラスの構造は液体と同じように乱雑なままであり、結晶の周期構造のような明確な秩序を有しない。この事情は、ガラス転移を特徴づけること、ガラスの内部の構造欠陥（の類似物）を抽出することを困難にしており、ガラスの理解および工学的応用に大きな障害となっている。ガラス科学において中心的な課題の一つは、構造的特徴のない中でどのように構造緩和するか、すなわちどの空間的領域にある原子（粒子）が時間とともに構造的変化を示すかを

予測することだが、この困難の中、2010 年代半ば頃から、ガラスに潜む特徴的構造を機械学習によって自動的に抽出する教師なし学習手法、ガラスの構造と動力学を結びつける教師つき学習手法が開発され、従来よりも遥かに高い構造抽出性能・予測性能を達成するに至った。

本研究課題では、これら一連のアプローチの中でも近年の発展を代表するグラフニューラルネットワーク (GNN) を用いたアプローチを軸に、深層学習モデルを活用したガラスダイナミクスの予測器の高度化や、ガラスシミュレーションを代理するような生成モデリングを開発することを目的としている。

3 当拠点公募型研究として実施した意義

本課題は、代表者の芝および副代表者の下川辺が 2023 年にガラス動力学の教師つき学習のための GNN モデル “BOnd TArgeting Network (BOTAN)” を開発・発表した (業績 [2]) ことを契機に、この GNN の利用発展と新たなモデリング、シミュレーションとの融合を意図して実施しているものであり、3 年度目となる 2025 年度をもって最終年度となる。

もともと、ガラスのダイナミクスに対する機械学習・深層学習手法の利用の研究としては、2020 年 Google DeepMind のグループによってガラスのダイナミクスを予測させる GNN が他の機械学習・深層学習モデルよりも格段に高精度な粒子運動分布の予測を行うことができることを示した研究がある (V. Bapst et al., Nature Physics)。DeepMind は従来の機械学習モデルよりも大幅にデータ量に対する要求が強い GNN のモデルを提示にするに際し、巨大なステップ数にわたる直接分子動力学シミュレーションの軌道データを提出することによってその訓練を可能にしていた。

BOTAN はこのモデルに引き続き、モデルの予測精度を大幅に更新し、かつ改良されたデータセットを提出したものである。これは、ガラス物理のシミュレーションにおいて長い経験を持つ物性物理分野の研究者である芝が、高性能計算分野のリソースを有効に活用した大量のデータ創出をすることなくして不可能な成果である。特に、GNN の研究で使用したデータセットは、3 年間の課題の最大の成果とも言える、国際的なベンチマーキングデータセット “GlassBench” に結実している (2024 年度課題 jh2400** の報告参照) とともに、今年度には BOTAN の汎化性能の研究のための改良モデル T-BOTAN を川崎・別所とともに開発するに至った。また、生成モデルの検討については 3 年間の間の研究で顕著な進展をもたらすには至っていないが、芝が研究代表者として実施する科学研究費補助金のプロジェクトに引き継ぐ形で 2026 年度以降も継続して進めている。すでに得られた準備状況の進展に際しては、データサイエンスの専門家である鈴木・華井、またデータ駆動高性能計算を実施している下川辺・朝比というメンバー構成により、有機的な進展が得られたものである。

以上のように、本課題は緊密な、特定の物理分野のシミュレーション手法に強みを持った研究者と、高性能計算のリソースをシミュレーション・データサイエンスの両面から扱うことが可能な学際的研究体制を持っていることによって、成果に結びついたものということができる。またさらに特筆に値することとして、川崎は最終年度となる 2025 年度からは、JHPCN 構成機関である阪大 D3 センターのスタッフ (准教授) として着任している。

4 前年度までに得られた研究成果の概要

BOTAN 論文 (Shiba *et al.*, 2023) とともに芝・下川辺らが公開に供した、3次元 Kob-Andersen Lennard-Jones ガラス形成液体基準データセットとして用い、分野で提案された主要機械学習モデルの持つ予測性能、転移学習性能などを整理し、データベース化した。結果は、一般の研究者の検証や利用に供することができるように、Zenodo リポジトリ^{*1}上にて公開された。

前項の作業の過程で、「固有構造」(Inherent Structure) と通称される状態を学習の入力として使用することで、予測性能が多くのモデルで向上することが明らかとなった。ガラスの遅い緩和とは、概念的には多粒子座標の関数であるエネルギー地形における多数の準安定状態の極小を徐々に降っていくダイナミクスとしてとらえることができる。通常、有限温度ガラスでは、熱ゆらぎの影響があるため各粒子が運動中心の周辺で熱運動しているため極小値のところからわずかにずれることになる。鞍点(準安定状態)から鞍点へと長時間をかけて遷移するガラスのダイナミクスと対応づけるには、熱振動下でのダイナミクス準安定状態=エネルギー極小を満たす配置の方が「捉えやすい」入力であることが示唆される。

なお、熱的な粒子配置の状態から(最急降下法などによる準安定構造へのクエンチを行わずに)、直接粒子の運動度を予測するタスクにおいては、BOTAN は他の(教師つき学習の)モデルを凌駕してトップの予測性能に位置していることも明らかにされた。BOTAN はグラフの辺(edge)上で2粒子間の距離変化を学習す

ることにより、隣接粒子間で生じる弾性歪みの度合いが予測可能なモデルに訓練され、その情報が、学習モデルの内部で背後に隠された固有構造を「推定する」のに寄与していると考えられる。

動的多体相関を学習できる GNN モデルの開発今年度課題においては、これまでの GNN が部分的に物理的に不自然な予測を与えることに着目してモデルの改良を実施している。レビュー原稿 [1] の執筆過程で、BOTAN を含む GNN は損失関数に単純な平均自乗誤差または平均絶対誤差を用いているために、入力された各粒子の局所構造と動力学の間に過剰なフィッティングをかけてしまうという問題が認識されるに至った。実際には粒子の構造再配置が活発な領域は空間的に広がりを持っているのであるが、GNN においては近接する粒子同士の運動活性の相関を無視して、個別の粒子に合わせすぎた予測をしばしば行ってしまう。昨年度から、課題参加者の別所が主体となって、動力学の協調性の再現ができないことに対するペナルティを表現した GlassMLP モデル (G. Jung *et al.*, 2023) の損失関数を BOTAN に対して実装し、検証を進めてきた。この際、予測結果を単にピアソン係数のような統計尺度で見のではなく、予測結果から得られる構造関数や、時間を経たあとの粒子の運動距離がどのように分布しているか(動的不均一)、などといった物理量を通じても比較し、その予測の妥当性を検討した。その結果は粒子の運動距離の分布の確率密度分布や、動的不均一性を特徴付ける量を見た時に、分布が学習元のデータセット(シミュレーションデータ)により近いものに改善した。

^{*1} <https://zenodo.org/records/10118191>

5 今年度の研究成果の詳細

今年度は、前年度までに得られた「動的多体相関を学習できる GNN モデル」を基盤として、GNN をさまざまな条件で訓練した場合に獲得される予測性能を、より詳細に理解するための研究を実施した。

BOTAN は、頂点に個々の粒子、辺に近接粒子ペアを割り当て、ガラスを形成する粒子集合体をグラフとして表現した上で動作する GNN である。グラフに構造情報を特徴量として付与し、GNN への入力情報とするとともに、動力学情報をグラフ上の出力として割り当てることで予測を行う。今回、BOTAN を拡張し、「グローバル特徴量」を導入したモデル T-BOTAN を考案・実装した。T-BOTAN では、頂点特徴量からのメッセージパッシングを通じてグラフ全体の特徴を集約し、温度を推定する出力を持たせた。グローバル特徴量は、ノードに埋め込まれた特徴量からの情報を受け取る一方で、グローバル特徴量からノード特徴量への情報伝搬は行わない。すなわち、グローバル特徴量の出力は、グラフから抽出された特徴量をスカラー化して出力する役割にとどまる。ここで学習対象を温度とした場合、グローバル特徴量とそのデコーダーは、いわば「温度計」としての機能を獲得することになる。

構築した T-BOTAN に対し、BOTAN データセットを拡張し、より多数の温度をカバーする 3 次元 Kob-Andersen Lennard-Jones ガラス形成液体のデータセットを構築した。対象温度は、従来の BOTAN データセットで使用されていた $T = 0.44, 0.50, 0.56, 0.64$ に加え、新たに $T = 0.47, 0.80$ の 2 温度を追加した計 6 温度である。このデータセットは、Zenodo リポジトリとして整備・公開している。これら 6 温度のうち、可能な複数の 4 温度の組み合わせ

せでモデルを訓練し、訓練に用いなかった残り 2 温度のテストデータに対して推論を行った。その結果、以下の知見を得た。

1. T-BOTAN では、学習データに含まれない未知の温度 T_{new} に対しても、個々のスナップショットでは既習温度に近いダイナミクスを返す傾向が見られた。しかし、これらのアンサンブル平均を取ることで、 T_{new} において本来観測されるべきダイナミクスを正確に補完できることが明らかになった。

さらに、このようにして得られたダイナミクスの不均一度の空間分布を、波数空間において解析する 4 点構造因子により評価した。その結果、平均二乗誤差を損失関数に用いる従来の BOTAN で見られるような、系全体に対する運動度の過大評価や標準偏差の顕著なずれを伴わずに、ダイナミクスを評価できることがわかった。

この結果は、T-BOTAN のみならず、従来から提案している BOTAN においても、十分に多数の温度で訓練されたモデルが内挿によりダイナミクス情報を適切に補完できることを示している。すなわち、BOTAN 系列のモデルは、ピアソン係数のような統計的尺度による精度評価を超えたレベルで汎化性能を有しているといえる。また、これらの結果は、BOTAN が抱えていた予測精度の定量性に関する実用的課題を解決するものと位置づけられる。

2. 前項の結果と関連して、内挿が統計的に成功している場合であっても、内挿に用いる未学習温度が学習温度から大きく離れている場合には、より近い学習済み温度における学習内容に予測結果が引き寄せられる傾向がある。T-BOTAN は、温度予測を

行うためのグローバル埋め込みに潜在表現を有するアーキテクチャであることから、温度の潜在表現に関する距離行列（ヒートマップ）を作成し、この傾向を定量的に評価した。

この結果から、本質的に多温度条件へ対応可能なモデルへと改良するためには、(A)十分に多様な温度条件のデータを用いてモデルを訓練することに加え、(B)温度とダイナミクスとの関係を経験則として定式化し、そこから得られる拘束条件を損失関数に組み込むことなどが有効なアプローチとして考えられる。

3. 従来は BOTAN では、GNN への入力として有限温度の熱的粒子配置を使用していた。BOTAN の精度向上は、グラフの辺に対応する埋め込みを用いて近接粒子間の相対運動を学習することにより、静的配置における熱揺らぎによるずれと、熱揺らぎを取り除いた場合に観測されるべき Inherent Structure を切り分けられることによって得られたものと考えられる。

今回、GNN への入力として、(1)有限温度の熱的配置 (Thermal Structure, TS)、(2) (1) の配置を FIRE アルゴリズムによりゼロ温度へクエンチすることで得た Inherent Structure (IS)、(3) TS と IS の情報を組み合わせた入力 (TS+IS) の 3 パターンを用意し、グローバル特徴量のデコーダーに温度を推定させるよう T-BOTAN を訓練した。その結果、いずれの構造を入力とした場合でも、そのガラスがもともと TS として作成された温度を予測できるように、T-BOTAN を訓練できることが確認された。

異なる温度で作成されたガラスから、ゼロ温度へのクエンチ後に得られる IS 同士の

構造的差異は、有限温度の配置 (TS) 同士の差に比べると極めて小さい。したがって、T-BOTAN は、もとのガラス粒子配置が宿している熱力学的情報が指紋のように保存された構造を抽出することにも成功しているといえる。この結果は極めて興味深く、Inherent Structure に潜むダイナミクス支配因子の解明をさらに動機づけるものである。

以上得られた結果は、現在、原著論文として投稿、査読中である (arXiv:2603.13820)。

6 進捗状況の自己評価と今後の展望

2025 年度の研究計画は、以下の 3 項目に分けて実施した。

第一の項目である「(A) ガラスの動力学予測システムの本体の開発」については、固有構造を積極的に活用したモデリングへと GNN を改良し、動力学予測を温度軸方向に拡張することに成功した。最終的に、レギュラージャーナルへの論文投稿にも至っている。当初計画とは一部異なる方向での発展ではあったものの、研究目的に照らして重要な進展が得られたことから、十分な進展があった評価できる。

第二の項目であり、本研究課題の主要な目標の一つである「(B) 生成モデルによるガラス動力学のサンプリング」については、研究期間中に、連続正規化フローを用いた G. Jung らのモデル (Mach. Learn.: Sci. Technol., 2024) に対して我々独自のデータセットを構築し、評価を実施した。この成果については、分子シミュレーション討論会 (島根) においてポスター発表を行った。さらに、拡散モデルの実装も進めることができた。一方で、JHPCN 研究期間内に一つの手法体系として完成させる段階には至

らなかったため、進捗度は当初想定より低い。しかし、2026 年度以降、新たに科学研究費補助金 基盤研究 (C) 課題「ガラスダイナミクスの超長時間サンプリングに向けた構造遷移生成モデルの開発」が開始される予定であり、本課題を通じてその研究基盤を形成することができた。

第三の項目である「(C) 様々な高性能計算用加速デバイスへの GNN の展開」については、当初、GNN を AI 専用機と呼ばれる機種で動作させることを検討していた。しかし、そこで求められる開発要素が、本研究体制で実現可能な範囲をやや超えていることが判明したため、GNN の学習を開発対象とする方針は中止した。その後、2025 年 11 月頃にテーマ設定を再検討し、ガラスシミュレーションそのものに AI 専用機を利用する方向へと方針を転換した。現在は、この方向に基づきコードの試作を継続している。

以上を総合すると、本課題では、GNN に基づくガラス動力学予測モデルの高度化という中核的目標について、論文投稿に至る明確な成果を得ることができた。他のテーマに期間内に完成形を示すには至らなかったものの、今後の研究課題へと接続する基盤を形成することができた。