

結晶構造の類似性に基づいた物質探索手法の研究

草場 穫 統計科学専攻 博士課程4年

1. 研究概要

計算機による無機材料探索の問題において前提となる最も基本的なタスクは、特定の化学組成の結晶構造を予測することである。結晶構造予測手法として、遺伝的アルゴリズム(Glass, 2006)、粒子群最適化(Wang, 2012)、ベイズ最適化(Yamashita, 2018)などを用いた様々な手法が提案されている(Terayama, 2018)が、いずれも数多くの第一原理計算を要求するため、非常に多くの計算コストを必要とする。そこで本研究では、ある化学組成が与えられた時の結晶構造の予測問題を、特定の結晶構造のグループと同じ結晶構造を持つか否かを判定する問題に置き換えることで、コストが高い計算処理を回避し、高速かつ実用的な無機材料探索手法の構築を試みる。

2. 背景と課題

ある無機化合物の物性を第一原理計算によって求める上で、化学組成に加えてその組成に対応した最安定構造を知る必要があるため、結晶構造予測の問題は無機化合物探索のワークフロー全体(図1)の中で最も基本となるタスクであり、この処理が遅いとボトルネックになってしまう。最安定な結晶構造は最も低いエネルギーを持つ構造であるため、結晶構造予測のタスクはエネルギーの最小化問題と同一である。また、ある結晶構造に対するエネルギーは密度汎関数理論に基づく第一原理計算によって正確に計算することができる。従って前節で挙げたように、この結晶構造予測の問題を未知関数の最適化問題として様々な手法が提案されている。これらの手法は強力である反面、数多くの第一原理計算を必要とするため計算コストが非常に高い。

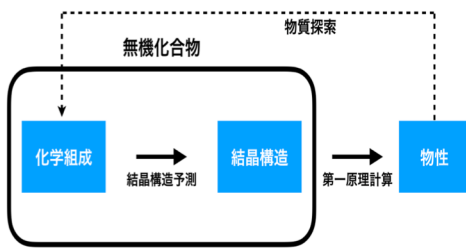


図1 無機化合物探索問題のワークフロー

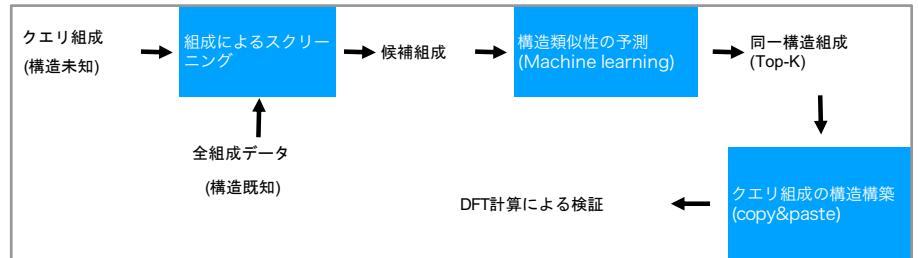


図2 提案する手法のワークフロー

3. 提案する手法

本研究では、ある新しい化学組成(クエリ組成と呼ぶことにする)の結晶構造が知りたい時、以下の手順によってその結晶構造の予測を試みる。まず、結晶構造が既知である大量の化学組成データが手元にある状態を仮定する。次にそのデータによって訓練された、結晶構造の予測モデルを用意する。予測モデルとして、データ間の類似度を学習する手法である距離学習を採用した。ここでは、2つの異なる化学組成が与えられた時、それらの結晶構造が同一か否かを判別するようにモデルを学習させている。いくつかの距離学習手法を試したが、DNN(Deep Neural Network)が最も良い精度を与えた。この訓練済みモデルにクエリ組成と構造既知である手中の組成を入力し、同一構造である可能性が高いと推定される順に構造既知組成を並び替えることで、クエリ組成の結晶構造と類似性が高いと推定される結晶構造を列挙することが可能である。そして、構造既知組成の元素をクエリ組成のものに置き換えることでクエリ組成の結晶構造予測を行うことが出来る。最後にその構造を初期構造として第一原理計算を行い、真に安定構造であるか検証を行う。提案する手法のワークフローを図2に示す。

4. 事前スクリーニングの重要性と結晶構造予測結果

本研究では、ある新しい化学組成訓練済みモデルによって構造の同一性を判別する時、正例:負例=1:1ならばそれなりに良い精度で判別できることが分かった。しかしクエリ組成が与えられた際、全組成データに含まれる正例(構造同一)の割合は非常に少なく、大量の偽陽性に埋もれてしまう問題が起こる。よって負例をあらかじめ排除することで偽陽性を減らす必要がある。これを実行する手法として、クエリ組成と同じ元素の組成比を持つ構造既知組成だけを候補組成として選択するという事前スクリーニングを適用した。また、この過程は元素種の置き換えだけでクエリ組成の構造予測ができることを保証するので実用上極めて重要である。

上記の手続きに基づいて、事前に用意された40個のクエリ組成に対する結晶構造予測を行った。判別モデルとしてはDNN(精度をあげるためアンサンブル化している)を採用し、上述の事前スクリーニングの後に、クエリ組成ごとに上位30個までの結晶構造の提案を行った。図3に上位N個ごとの予測精度を示す。図にあるように上位30構造まで提案すると82.5%、上位10構造まででも80%のクエリ組成に対して、真の構造と類似する構造が予測できたことが分かった。尚、ここで構造が類似であるとはLocal order parameters(Zimmermann, 2020)によって計量された構造非類似度が0.75以下であると定義している。

5. 第一原理計算による検証

上述の予測において、40個のクエリ組成ごとに10個までの結晶構造の提案(予測)を行い、第一原理計算ソフトであるquantum espressoによって提案された結晶構造の検証を行った。全370件(各クエリに対して必ずしも10個の提案構造が存在するとは限らないため400ではない)の計算に対して、149件が収束条件を満たした。また、組成ごとに見ると40個中31個(77.5%)の組成に対して1つ以上の収束条件を満たす構造を提案することが出来た。さらに、第一原理計算によって構造最適化された提案構造と、40個の化学組成の真の最安定構造を比較した結果、40個中12個(30%)の組成に対して1つ以上の真の最安定構造と同一な構造を提案することが出来た。図4に結晶構造予測結果のまとめを示す。

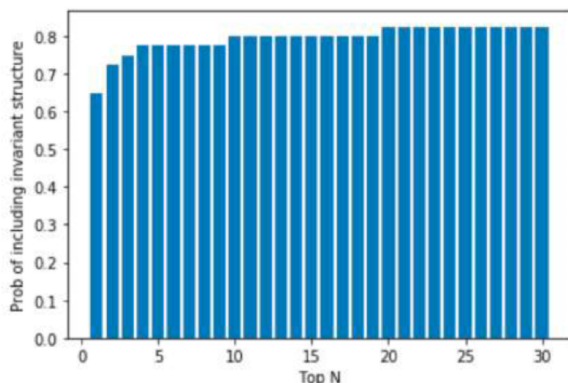


図3 上位N個ごとの予測精度

pretty formula	comment	ZnSb	熱電材料
C	多形	CoSb3	熱電材料
Si	多形 LAQA	LiBF4	電池添加剤
GaAs	半導体 LAQA	Y2Co17	LAQA
ZnO	多形	GeH4	USPEX
BN	多形	CsPbI3	ペロブスカイト
LiCoO2	電池、層状	NaCaAlPO5F2	元素数が多い
Bi2Te3	熱電、長いユニットセル	LiFePO4	電池、正極
Ba(FeAs)2	超伝導体	Cu12Sb4S13	熱電材料
SiO2	多形	MgB7	USPEX
VO2	多形	Li3PS4	電池、0次元
La2CuO4	超伝導	Cd3As2	ディラック半金属
LiPF6	電池添加剤	Li4Ti5O12	電池負極
Al2O3	多形 LAQA	Ba2CaSi4(BO7)2	元素数が多い
SrTiO3	ペロブスカイト	Ag8GeS6	熱電, argyrodite
CaCO3	CALYPSO	Nd2Fe14B	磁性
TiO2	多形	Y3Al5O12	蛍光体 (YAG)
ZrO2	多形	Ca14MnSb11	構造が複雑
ZrTe5	ディラック半金属	# # 色の説明 # #	
V2O5	多形、触媒		
Si3N4	セラミックス		→同一構造あり(12個)
Fe3O4	磁性		→類似構造なし(8個)
Mn(FeO2)2	スピネルフェライト		→類似構造あり、同一構造なし(20個)

図4 40個のクエリ組成に対する結晶構造予測結果まとめ