

修士論文

機械学習と第一原理計算に基づく 大規模強相関係物質の誘電特性予測

令和3年度

三重大学大学院 工学研究科

博士前期課程 物理工学専攻

ナノサイエンス・ナノテクノロジー領域

松本 崇宏

三重大学大学院 工学研究科

目次

第1章 序章	4
1.1 マテリアルズインフォマティクス	4
1.2 光通信環境の発展とイットリウム鉄ガーネットの役割・課題	6
1.3 イットリウム鉄ガーネット	7
1.4 第一原理計算における光学物性と $U_{\text{eff}}, 1/\tau$ の関係と問題点	8
1.5 本研究で開発する手法と関連研究	11
1.6 本研究の目的と論文の構成	13
第2章 全電子フルポテンシャル線形補強平面波法とオンサイトクーロン相互作用	15
2.1 はじめに	15
2.2 FLAPW 法	17
2.3 +U 法による電子強相関効果補正項	22
2.4 +U 法の物理的解釈	24
2.5 第一原理計算に基づく誘電関数と磁気光学カー効果の計算	26
2.6 おわりに	29
第3章 ニューラルネットワークによる機械学習	30
3.1 はじめに	30
3.2 データの前処理と訓練・テストデータの分割	32
3.3 データ拡張による予測モデル性能向上	34
3.4 ニューラルネットワークの訓練方法	35
3.5 評価方法	41
3.6 計算フロー	43
3.7 ニューラルネットワークを用いた $U_{\text{eff}}, 1/\tau$ 予測手法	45
3.8 おわりに	46
第4章 NIO における機械学習手法の妥当性の検証	47
4.1 はじめに	47
4.2 第一原理計算による誘電関数データの収集	48

4.3 機械学習に用いるデータ選択	49
4.4 学習・推論モデル	51
4.5 第一原理計算との比較	55
4.6 まとめ	57
第5章 YIGにおけるパラメータ予測とデータ拡張による性能向上	58
5.1 はじめに	58
5.2 第一原理データ収集	59
5.3 学習・推論モデル	60
5.4 ガウシアンノイズによるデータ生成と拡張パラメータの選定	63
5.5 ガウシアンノイズによるデータ拡張と荷重減衰による正則化の比較	67
5.6 推定したパラメータを用いた第一原理の光学物性と実験結果の比較	69
5.7 まとめ	72
6. 結論	73
参考文献	75
謝辞	79

第 1 章 序章

1.1 マテリアルズインフォマティクス

従来の材料デザインは、図 1.1 のように材料の格子定数、磁気構造や外部電場を仮定し、密度汎関数理論などの理論解析、実験を行うことで電子構造、各種物性の予測を行い、複数回の試行錯誤を経て所望の物性をもつ材料を探索していた。しかしこの手法では、材料デザインは研究者・組織の経験と勘を頼ることになり、試行時間・費用両面から莫大な労力が必要であった。また、革新的な性質を持った材料は偶発的に見つかることも珍しくなく、材料デザインは属人的な側面が大きかった。そこで機械学習やデータサイエンスの手法により所望の物性をもつ材料をデザインするマテリアルズインフォマティクス(Materials Informatics, MI)が重要視され始めた。2011 年、米国で世界的な MI ブームの先駆けとなる Materials Genome Initiative (MGI) がオバマ大統領（当時）の主導でスタートした。その後世界中で MI に関する研究開発プロジェクトが開始しており、2015 年には欧州では Novel Materials Discovery (NOMAD) が、2014 年にはスイス独自の取組として Materials Revolution Computational Design and Discovery of Novel Materials (MARVEL) が開始されている。そして MI が著名になった契機は 2015 年のリチウムイオン電池の構造最適化の論文[1]である。この論文では、固体リチウムイオン電池を、当時広く流通していた液体電解質による二次電池に匹敵する程度まで電気伝導度を上昇させるため、機械学習手法を用いて最適な結晶構造を探索に成功した。この研究では実験を一切することなく第一原理計算とデータサイエンスのみで完結している点で、当時、実験によるトライ&エラーで探索していたコミュニティに対して大きな衝撃を与えた。以来、MI を用いた研究事例は年を経るごとに増加しており、今日では材料科学の一角を占めるようになった。次節では本研究の主題である強相関電子系における課題について述べ、MI 的手法による解決法について記述する。

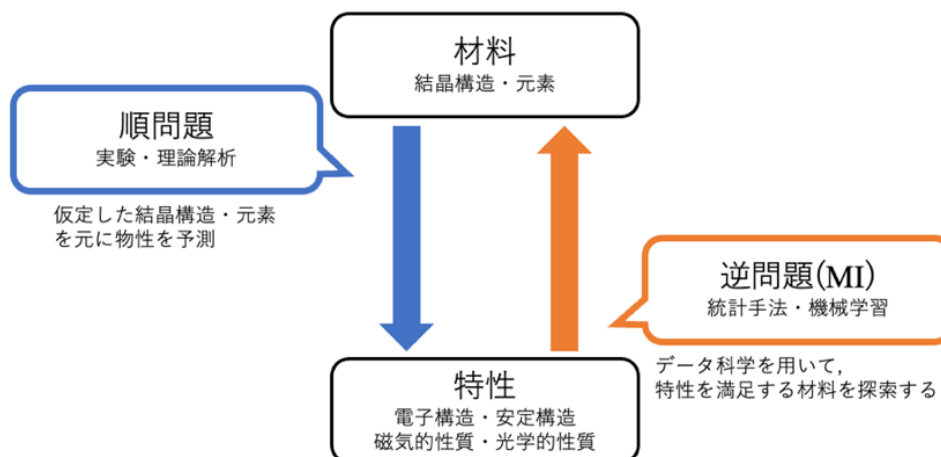


図 1.1 実験・理論解析と MI を用いた材料デザイン手法の概念の比較. 材料の情報を仮定してから特性を観測・予測する手法を順問題とし、逆方向に予測を行う MI を逆問題と定義した.

1.2 光通信環境の発展とイットリウム鉄ガーネットの役割・課題

2022 年現在, インターネットをはじめとする高度な情報処理が行われるようになり, 年々扱われる情報量は指数関数的に増加している. 総務省によると 2019 年 11 月時点で日本国内での 1 秒当たりのデータ通信量は約 12.7 Tbps[2]あり, 今後も, 動画配信・SNS などのオンラインコンテンツの増加やサービスの向上により世の中に存在する情報量はますます増加すると考えられている. このような大規模データのサーバー, 個人間通信を支えるのは光ファイバーを主とした光回線であり, 近年ますます需要が大きくなっている. その中でも外部磁場によって光を一方向に整流する光アイソレータや, 旋光を制御する光サーキュレータは特に重要であり, これらの磁気光学(Magneto-Optical; MO)デバイスの材料候補としてイットリウム鉄ガーネット (YIG) が注目を集めてきた. YIG は多くの実験にて大きなファラデー効果や磁気光学カー効果を示すことを報告されており, 先述の MO デバイス設計のため YIG による MO 学効果の理論的の解明が急がれている.

YIG の理論的理解のためには MO 効果の基礎物性である誘電関数を, 高効率かつ高精度に理解・解明することが重要である. しかし密度汎関数理論 (DFT) に基づく第一原理計算の定量化には, いくつか課題が残されているのが現状である. 例えば, 大規模強相関系に適用可能なハバードモデル (DFT+ U 法) [3]のモデルパラメータである有効オンサイトクーロン相互作用 (U_{eff}) の値や, バンド間遷移に伴う緩和時間 ($1/\tau$) の値を実験結果に最適化する必要がある. これらの任意パラメータの信頼性の向上が, 理論計算に基づく MO デバイス設計を加速させると考えられるが, これらパラメータは誘電関数の形状に対して複雑に作用し合い, わずかな変化に対して容易にその形状を変化させる. そのためこれらの変化を正確に予測できる非線形回帰が重要な鍵を握っている. そこで本研究では誘電関数から U_{eff} と $1/\tau$ を MI 的手法を応用して最適化する手法を開発し, この問題の解決を目指した.

1.3 イットリウム鉄ガーネット

イットリウム鉄ガーネット(Yttrium iron garnet: YIG)は、中心対称性を有する立方晶系であり、空間群 $Ia\bar{3}d$ に属する。YIG の結晶構造は Gellar と Gilleo[4]により結晶構造が解明された。YIG はユニットセル内には 160 個の原子を含み、化学式は $Y_3Fe_5O_{12}$ と表される。ユニットセル内において、Y イオンが十二面体配位、5 個の Fe イオンのうち 2 個が八面体配位、3 個が四面体配位を占有し、その周りに O イオンが配位している。その中でも八面体配位と四面体配位を占める Fe イオンがそれぞれ反強磁性的に結合しており、全体としてフェリ磁性を有する。図 1.2 にその結晶構造を図示した。

YIG は発見以後、1950 年代から 1960 年代にかけて、その磁気・光学特性に注目が集まり、実験と理論の両面で多くの研究が報告された[5-10]。またデバイス応用としては、磁気減衰の低さやバンドギャップの大きさなどからフィルター[11, 12]やセンサー[13, 14]などのさまざまなマイクロ波デバイスとして注目された。さらに可視光から近赤外波長領域における低い光吸収や磁気光学ファラデー効果、カー効果などを組み合わせた磁気光学アイソレータ[15, 16]のような通信機器への応用も期待され、多くの研究が報告された。現在もビスマスやセリウムを YIG にドープさせることにより、磁気光学効果の増大が報告され注目を集めている[17, 18]。

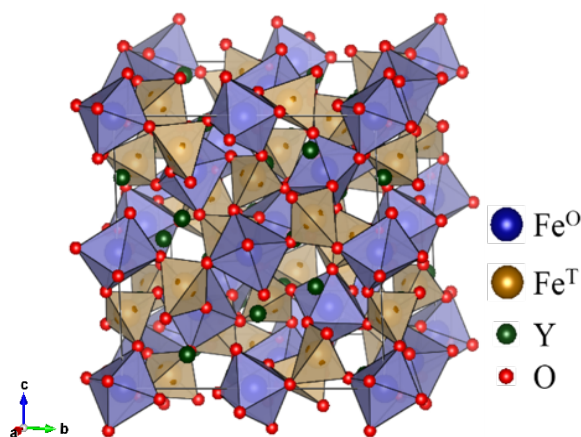


図 1.2 立方晶としての YIG のユニットセル。緑色がイットリウムを示し、紫色と茶色の構造は酸素（赤色）による八面体格子と四面体格子を示す。またそれぞれの構造の内部にそれぞれ鉄イオン Fe^0 (青色)と Fe^T (金色)が存在する。

1.4 第一原理計算における光学物性と $U_{\text{eff}}, 1/\tau$ の関係と問題点

第一原理計算をはじめとした理論計算においては, YIG の電子構造やその基礎物性に関する研究報告すらほとんどないのが現状である. これは YIG の結晶構造が複雑であり, 取り扱う原子数が膨大であることや第一原理計算における電子強相関効果の取り扱いが不十分であることなどが考えられる. 本論文では密度汎関数理論(Density functional theory; DFT)に基づく第一原理計算により, 材料の電子状態解析を行う. 一般に密度汎関数理論においては Kohn-Sham 方程式を解くことで電子の基底状態を数値的に解析する. この方程式を解く際, 1 電子近似に基づく有効ポテンシャルを設定し, 変分原理に従い全エネルギーが最も低くなる状態を自己無撞着場(Self-Consistent field: SCF)により探索する. しかしこの密度汎関数理論による基底状態探索にはいくつか問題を抱えている. 例えば, 本研究で用いる強い電子相関をもつ多電子系を, 一電子 (平均場) 近似では適切に取り扱うことができないことである. 金属や半導体などの遍歴電子系は図 1.3(a)にあるように電子が他の電子から感じるクーロン相互作用が弱い為に結果として系の広範囲にわたってその電子が振る舞うことができる系である. そのため DFT における相関交換項は価電子を空間渡って平均的に分布すると近似する平均場近似を適用した局所密度近似(Local Density Approximation; LDA)[19]や一般化勾配近似(Generalized Gradient Approximation; GGA)[20]を用いて近似する. しかし図 1.3(b)のように強相関電子系の価電子である d もしくは f 軌道電子が, 電子間の相関が強くなり原子核近傍に局在するため, LDA や GGA など遍歴電子系の近似手法では近似しきれないことである. そのため遍歴電子系の相関交換項にオンサイトクーロン相互作用 U による計算補正を行う. 代表的な例として挙げられる遷移金属酸化物において, 局所スピ密度近似に基づいた計算では, NiO や MnO の基底状態における磁気構造を再現し, 極めて小さいながらバンドギャップを持つ絶縁体であると予測するが, CoO や FeO においては金属であると予測してしまう[21, 22]ということが報告されている. またこのような強相関電子系においては, マルチバンドの Hubbard 模型や Anderson 格子模型[23-26]によって, 電子状態がうまく記述されることが既に報告されており, これらの模型における本質的な仮定である強い相関を持つ d , f 電子がオンサイトの準原子相互作用を受けることを適切に取り入れる必要があるといえる. 図 1.4(a), (b)は金属と絶縁体の状態遷移と各状態での状態密度図を模式的に示したもので, 遍歴電子系のようにバンド幅が広がることで金属的な性質を示し, クーロン相互作用が大きくなることで絶縁体的な性質を示す. このように金属-絶縁体間の状態遷移にクーロン相互作用力が重要な役割を担っている. したがってこのような電子間の多体効果を適切に考慮し, 補正することが必要である. また U は光学応答にも影響を及ぼす. ここで有効オンサイトクーロン相互作用 U_{eff} を U と交換相互作用 J の差($U_{\text{eff}} = U - J$)にて定義し, 図 1.5(a)に YIG の誘電関数に対する 2 種類の異なる配位子 (八面体配位, 四面体配位) における Fe イオンに対する有効オンサイトクーロン相互作用である $U_{\text{eff}}^O, U_{\text{eff}}^T$ の依存性を示す. 図 1.5 から, 誘電関数などの光学応答において U_{eff} はバンド間遷移における占有軌道と非占有軌道のエネルギー差に直結しており, 誘電関数では主にピーク

の位置に対して影響する傾向がある．一般に U_{eff} の増大に伴いバンド幅が広がるため誘電関数のピーク位置は高い光子エネルギーへシフトする．しかし YIG の場合 2 つの U_{eff} が誘電関数のピーク位置に寄与しており， U_{eff} と誘電関数のピーク位置は大雑把な正の相関が見られるものの， U_{eff} と $1/\tau$ の変化によってスペクトル形状が複雑に変化する．また誘電関数のバンド間遷移における緩和時定数である $1/\tau$ は図 1.5(b)のようにピーク周辺の分散に寄与しており，誘電関数の実験を再現するにはこれらの特徴の変化を捉える最適化が必要となる．

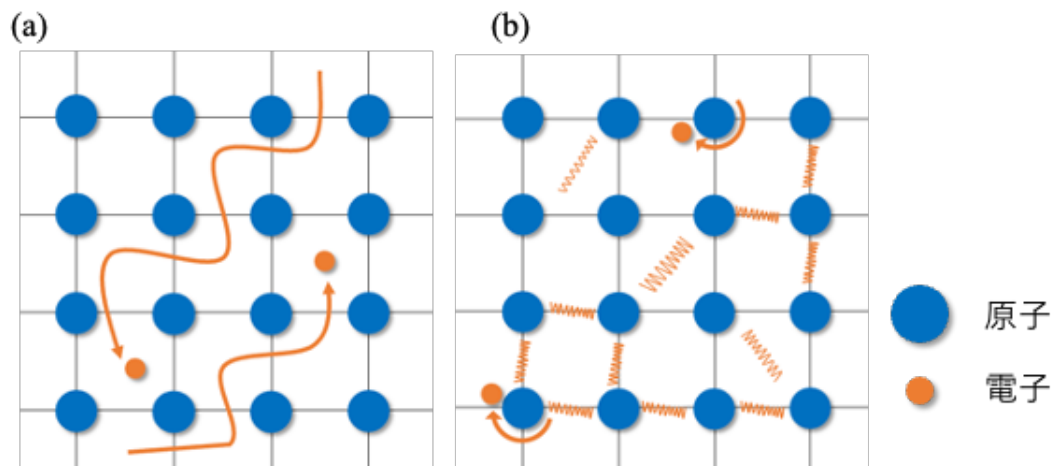


図 1.3 (a)遍歴電子系及び(b)強相関電子系における電子の振る舞いの概念図．
遍歴電子系では，電子は自由電子のように系全体を動き回るのに対し，強相関電子系では電子間に強いクーロン相互作用が働くため電子が原子核付近に局在する．

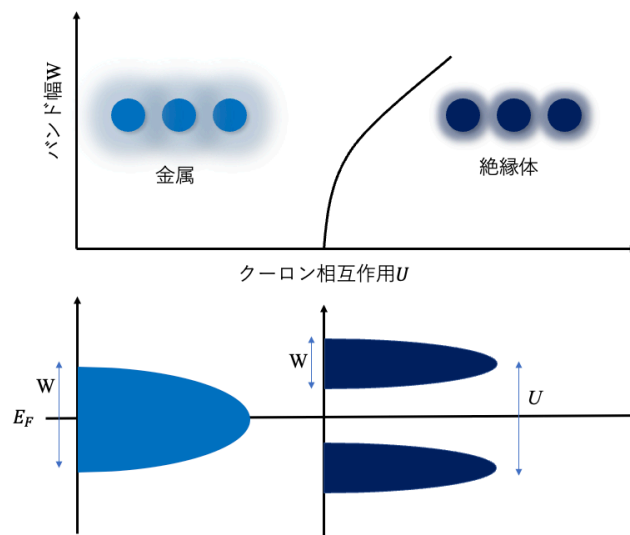


図 1.4 (a)金属-絶縁体間の状態遷移の模式図(b)金属と絶縁体の状態密度の模式図（左が金属,右が絶縁体の模式的な状態密度）．バンド幅 W とクーロン相互作用 U により状態の遷移を表している．

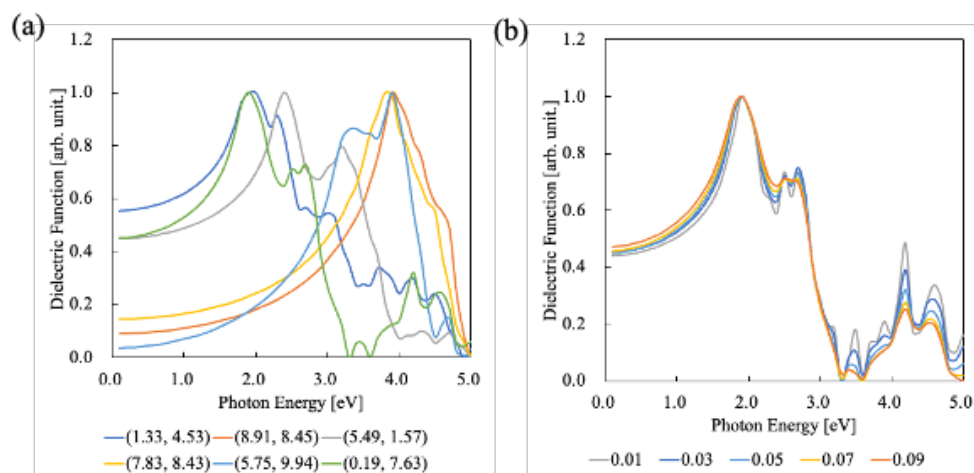


図 1.5 (a)第一原理計算より得られる誘電関数実部の U_{eff} 依存性及び(b)バンド間遷移に伴う緩和時定数 $1/\tau$ の依存性. (a)では $1/\tau$ を 0.05eV に固定し, $(U_{\text{eff}}^o, U_{\text{eff}}^T)$ (単位: eV)依存性, (b)では $(U_{\text{eff}}^o, U_{\text{eff}}^T) = (0.19, 7.63)$ の時の $1/\tau[\text{eV}]$ 依存性を描写.

1.5 本研究で開発する手法と関連研究

1.4 では第一原理計算における強相関効果の取り扱いと光学応答への影響について述べた。複数のパラメータが同時に複雑に誘電関数の形状に影響することから、これらの最適化には非線形かつ複数のパラメータを同時に予測できる手法が必要となる。そこで本研究では誘電関数のスペクトルから U_{eff} と $1/\tau$ をニューラルネットワークによって予測する手法を構築し、実験から観測された誘電関数を入力することでこれらパラメータの最適値を予測する経験的手法を構築する。これは通常の U_{eff} と $1/\tau$ を定めて実施する第一原理計算と、そこから得られる電子構造から誘電関数 ϵ を計算する一連の過程を

$$\epsilon = f(U_{\text{eff}}, 1/\tau) \quad (1.1)$$

とした時、誘電関数からパラメータを計算するという逆方向の演算する過程

$$U_{\text{eff}}, 1/\tau = f^{-1}(\epsilon) \quad (1.2)$$

をニューラルネットワークによる機械学習で近似し、実験から得られた ϵ を用いて U_{eff} と $1/\tau$ を推定する。

本手法の肝要は U_{eff} などのパラメータと誘電関数などの物性の関係を数理モデルで近似することで最適化を図ることにあるが、これらについて様々な類似した研究がなされており、強相関電子系の研究テーマとして関心を集めている分野であるといえる。まず YIG における U_{eff} を推定するために「線形回帰を用いてフィッティングする方法」が挙げられる。これは第一原理計算から得られるバンドギャップと U_{eff} の間について線形回帰し、実験から観測されるバンドギャップと良い一致を示したものを採用する手法であり一定の成功を収めている[27, 28]。また Optical Magnetic Circular Dichroism (OMCD) スペクトルより得られたバンドギャップにフィッティングを行う試み[29]も行われている。しかしこれらの手法によって推定された値を用いて計算された誘電関数や OMCD スペクトルなど光学応答は依然第一原理計算と実験の間には定量的な一致には至っていないのが現状である。

他にも「数理最適化を用いてフィッティングする方法」も挙げられる。これは NiO の Ni d 軌道電子の U_{eff} に対し数理最適化手法であるベイズ最適化を用いて推定した手法[30]であり、定性的な NiO のバンド構造の最適化に成功した。この手法は HSE 方程式[31]から得られたバンド構造に対し PAW[32-35]基底における第一原理計算の U_{eff} を最適化しており、NiO では線形応答理論[36]による U_{eff} の導出と比較して短い時間で算出することが可能であり、計算コストの面で効果的である。しかし本手法は逐次的に U_{eff} の候補となる値を探索しており、1 回の電子構造計算が大きくなるとベイズ最適化による探索が収束までに膨大な時間を要するという困難を抱えている。またこの手法はフィッティング対象となる HSE 方程式から得られたバンド構造が必要となるため、本研究で目指す実験に対する最適化ではなく理論計算に対する最適化であるという側面も持つ。

以上の観点から、本手法は前述の線形回帰手法を YIG が持つ複数の U_{eff} を予測できるような多値予測が可能なニューラルネットワークによって拡張し、 $1/\tau$ も含めることで誘電関数から複数のパラメータを同時に予測する方法である。またニューラルネットワークはその内部を非線形化することが可能であるため、誘電関数に対する U_{eff} , $1/\tau$ の依存性のような複雑な関係を表現することが可能である。また $(U_{\text{eff}}, 1/\tau)$ -誘電関数データセットを予め構築してから回帰を行うため電子構造計算を並行して行うことができ、最適化を一定時間で収めることができるという利点も存在する。

1.6 本研究の目的と論文の構成

本論文では、大規模強相関電子系である YIG に対して、機械学習を用いて誘電関数から U_{eff} と $1/\tau$ の最適値を予測する手法の開発、少数データ時の予測性能改善手法の検討を目的とする。

本論文は全 6 章および付録から構成され、以下にその概要を示す。

第 1 章「序章」では強相関物質における U_{eff} と $1/\tau$ の取り巻く課題と本研究のコンセプトであるニューラルネットワークによる非線形予測手法について概要を説明した。また関連研究について本手法と簡潔に比較し、本研究の立ち位置について説明した。

第 2 章「全電子フルポテンシャル線形補強平面波法とオンサイトクーロン相互作用」では、本研究の重要となる全電子フルポテンシャル線形補強平面波(Full-potential Linearized Augmented Plane Wave: FLAPW)法に基づく第一原理計算手法について説明する。FLAPW 法は平面波基底を元にマフィンティン球による球面状の境界を仮定し、その外部を平面波、内部を補強平面波に置き換え線形化することで精度を維持したまま基底数を減らす手法である。また仮定するポテンシャルは非球対称であることから、より現実の系に近づけた計算を行うことができるバンド計算手法である。また本章では強相関材料に必要な有効オンサイトクーロン相互作用についても記述し、その物理的な取り扱いについても説明する。

第 3 章「ニューラルネットワークによる機械学習」では本研究で用いる非線形予測手法の基礎であるニューラルネットワークのアルゴリズムについて記述する。ニューラルネットワークによる予測モデルは大まかに訓練・テストの 2 つの段階を経て最適化される。まず訓練において、ランダムに初期化されたニューラルネットワークから誤差逆伝播法に基づく重み更新を行い、ニューラルネットワークの予測値と訓練の正解データを近づける手法について説明する。次にテストにおいて、訓練では用いなかったデータを用いてニューラルネットワークが未知データに対する性能（汎化性能）を評価する。この時先述の予測モデルの予測値と正解データの間の誤差のみならず、統計的手法を用いた評価も行い、そのモデルの汎化性能を評価する手法を説明する。

第 4 章「NiO における機械学習手法の妥当性の検証」では、本手法の妥当性を検証するため、典型的な強相関物質である NiO に対して実際にニューラルネットワークによる予測モデルを構築し、第一原理計算から得られたデータセットを用いて最適なモデルが作成可能か検証する。まず学習データ収集として第一原理計算による誘電関数の結果について説明し、次に誘電関数に対して前処理を行い、最も高性能な予測が期待できるデータを探索する。また訓練データを用いて予測モデルのハイパーパラメータの調整を行い、最も良い予測モデルについて検討する。最後に、実験で観測された誘電関数を訓練済み予測モデルに入力することで U_{eff} , $1/\tau$ を予測し、それらを用いた第一原理誘電関数と電子構造について実験と比較する。

第 5 章「YIG におけるパラメータ予測とデータ拡張による性能向上」ではパラメータを YIG の異なる 2 つ Fe サイトにおける U_{eff} と $1/\tau$ について NiO と同様に予測モデルを構築し、第一原理データセットを用いてパラメータを予測する。しかし本章では YIG で得られる第一原理データの数 NiO に比べて少数であり、少数データによる予測精度向上が必要となる。そのため訓練データに対してガウシアンノイズを付加することでオリジナルデータと似たデータを作りだし、訓練データの総数を増加させるデータ拡張を適用する。またデータ拡張のハイパーパラメータに対する予測モデルの性能変化についても検討する。最後に NiO と同じく実験から得られた誘電関数を訓練済み予測モデルに入力し、得られたパラメータによる第一原理誘電関数、電子構造について実験と比較した。また YIG がフェリ磁性であることから磁気光学カー効果についても実験と比較した。

第 6 章「結論」では本研究で得られた知見を統括して本論文の結びとする。

第 2 章 全電子フルポテンシャル線形補強平面波法 とオンサイトクーロン相互作用

2.1 はじめに

固体結晶中の電子状態を記述するために、1920 年代における量子力学の完成後ほどなくして、バンド理論の枠組みは作られた。1926 年に E. Schrödinger によって発表された Schrödinger 方程式[37]、1928 年に P. Dirac により Schrödinger 方程式に相対論を加えた形で発表された Dirac 方程式[38]、この二つを出発点として 1930 年前後には、F. Bloch の Bloch 関数[39]や E. Wigner と F. Seitz による Wigner-Seitz cell[40]など、その基礎理論に関わる研究は精力的に行われた。さらに 1930 年代から 1950 年代においては、J. C. Slater の補強平面波 (APW)法[41]、C. Herring の直交化された平面波(OPW)法[42]、J. Korringa, W. Kohn, N. Rostoker の名前をとって付けられた KKR 法[43, 44]、J. Slater と G. Koster の Tight-binding 法[45]、J. Phillips と L. Kleinman の擬ポテンシャル法[46]などの現在の第一原理バンド計算の基礎となる多くの手法が開発された。また 1964 年には P. Hohenberg と W. Kohn によって、Hohenberg-Kohn の定理[47]が提唱され、1965 年には、W. Kohn と L. J. Sham により実用的な Kohn-Sham 方程式[48]に定式化されることで、現在の密度汎関数理論[48]は完成した。これによりバンド理論によって解かれる一電子方程式に物理的解釈が加えられ、現在の第一原理バンド計算手法として、物性解析や材料設計の予測として広く用いられるようになった。さらに 1981 年には、E. Wimmer が提案したフルポテンシャル法[49]が注目を集めた。本論文でも、全電子フルポテンシャル線形化補強平面波法(Full-potential Linearized Augmented Plane Wave: FLAPW)法[49, 50]を採用する。FLAPW 法は価電子を含む全電子を取り扱うことが可能であり、物質の電子状態解析に用いられるすべての第一原理計算手法の中で、最も精度が高く、信頼できる結果を与える手法である。

しかしこのような計算手法の発展にもかかわらず、未だ第一原理計算による予測が困難な系は存在する。その代表例の一つと言えるのが、強相関電子系であり、金属原子の $d(f)$ 電子が原子核近傍に局在していることに起因して、電子間の相関効果の取り扱いが不十分であるために困難である。現在、その対処法として動的平均場近似[51]に基づいた計算手法や Hartree-Fock 法と密度汎関数理論の交換相関項を汎関数としたハイブリッド型汎関数[52]に基づいた計算手法が用いられている。しかし実装の面で、Hubbard 模型に基づき遮蔽された電子のオンサイトクーロン相互作用を導入する DFT+U 法が最も有効な手法の一つとなっている。

本章では, FLAPW 法の具体的な概念および表式についてその根底にある補強平面波 (APW) 法から概説する. また強相関電子系における電子間に働く交換相関項に対する近似的補正手法(+U 法)について述べる.

2.2 FLAPW 法

密度汎関数理論に基づく第一原理バンド計算手法は、ポテンシャル項や基底関数などの違いにより、本研究で用いる FLAPW 法をはじめ、KKR 法や LMTO 法など多岐にわたる。これらの選択は、Kohn-Sham 方程式を解く際の精度や計算コストの面で、非常に重要である。本節では、まず初めに FLAPW 法の根底にある APW 法、そしてこの APW 法の概説から出発し、この手法における数値計算上の問題として、非線形性について言及する。さらに、より現実的な原子のポテンシャルを仮定するためにマフィンティン近似を取り除いた全電子フルポテンシャル法について説明する。

ある原子の周りで解かれた価電子に対する動径関数は互いに直行しているので、この価電子の動径関数を用いて平面波基底関数を補強することができる。これが補強平面波法 (augmented Plane Wave : APW) 法[45]である。また、凝縮系において電子の感じるポテンシャルは原子核周辺で急激に変化し、原子核間では緩やかに変化する。そこで原子核を中心として適当な半径の球 (マフィンティン球という) を設定し、その内側と外側を別々に考える。これがマフィンティン (Muffin-Tin : MT) 近似と呼ばれるものである。その結晶ポテンシャル $V_{\text{MT}}(\mathbf{r})$ は、次式のように定義される。

$$V_{\text{MT}}(\mathbf{r}) = \begin{cases} V(r_a), & (r_a < S) \\ V_{\text{MTZ}}, & (r_a > S) \end{cases} \quad (2.1)$$

ここで、 \mathbf{r}_a は $\mathbf{r}_a = \mathbf{r} - \mathbf{R}_a$ で表される位置ベクトルである。ただし、 \mathbf{R}_a は原子の位置座標で、 S はその周りに仮定されたマフィンティン球の半径である。通常、一定値に仮定された格子間 (interstitial) 領域のポテンシャル V_{MTZ} がエネルギーの原点 (muffin-tin zero : MTZ) に選ばれる。このポテンシャル形状に対する過程が MT 近似である。MT 球内における基底関数は、展開係数を用いて次式で表せる。

$$\phi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\Omega} \sum_{l=0}^{\infty} \sum_{m=-l}^l i^l a_{lm}^k R_l(|\mathbf{R} - \mathbf{r}|) Y_{lm}(\widehat{\mathbf{R} - \mathbf{r}}) \quad (2.2)$$

ここで、 $Y_{lm}(\widehat{\mathbf{R} - \mathbf{r}})$ は球面調和関数、 $R_l(|\mathbf{R} - \mathbf{r}|)$ は動径関数である。また、波数ベクトルは $\mathbf{k}_i = \mathbf{k} + \mathbf{K}_i$ (\mathbf{k} : 第一ブリルアンゾーン内の波数ベクトル、 \mathbf{K}_i : 逆格子ベクトル) と定義する。次に、格子間領域における基底関数は、ポテンシャル V_{MTZ} が 0 により、

$$\phi_{\mathbf{k}}(r) = \frac{1}{\sqrt{\Omega}} e^{-i\mathbf{k}_i \cdot \mathbf{r}} \quad (2.3)$$

となる。MT 球内と格子間領域の基底関数は、MT 球面上で連続的な接続条件があり、これにより展開係数 $a_{lm}^{\mathbf{k}}$ が決定される。

$$a_{lm}^{\mathbf{k}} = 4\pi \frac{j_l(kS) Y_{lm}^*(\widehat{\mathbf{K}}_i)}{R_l(S; E)} \quad (2.4)$$

ここで球ベッセル関数 j_l は、 $j_l(x) \sim \frac{1}{x} \sin x - lx/2$ と定義される。次に APW 基底関数(式(2.3)および式(2.4))を用いて一電子波動関数を展開する。

$$\psi_i^{k_i}(\mathbf{r}) = \sum_{\mathbf{K}} \phi^{\mathbf{k}+\mathbf{K}}(\mathbf{r}) c_i^{\mathbf{k}+\mathbf{K}} \quad (2.5)$$

$\psi_i^{k_i}(\mathbf{r})$ が一電子ハミルトニアン H の固有ベクトルとなるには, APW 基底関数に関する行列式が

$$\det|(H - E)_{ij}| = 0 \quad (2.6)$$

となれば良い. ここで, MT 球面上で APW 基底関数は連続であるが, 一回微分が不連続であることに注意しなければならない. (2.6)式を解く場合, ある波数 \mathbf{k} を与えて, エネルギー軸 (E) について行列式を計算し, その 0 点を探す. しかし, APW 法の問題点は, 動径関数 $R_l(|\mathbf{r}|; E)$ がエネルギーに依存しているため, 行列要素がエネルギーについて非線形に変化することである. これは, 数値計算上の困難な点であり, あるエネルギー範囲に多数のエネルギー固有値が存在する場合, 例えば, 単位胞に多数の原子を含む化合物系の時に顕著になる.これを避けるため,式の行列式をエネルギーとして 1 次までを含む表式を与えたのが線形化法であり,これにより通常固有値問題と同様に行列の対角化により固有状態を求めることが可能になる.

線形化補強平面波法 (Linear Augmented Plane Wave : LAPW) 法[53]は, 最終的な方程式を対角化法により解くことが可能な永年方程式で記述できる. まず, 動径関数 $R_l(|\mathbf{r}|; E)$ をある特定のエネルギー値 E_l のまわりでエネルギー微分によるテイラー展開をし,その 1 次の項までを用いて表現する.

$$R_l(r; E) \approx R_l(r; E_l) + (E - E_l) \frac{dR_l(r; E)}{dE} \quad (2.7)$$

以降簡略化のため $\dot{R}_l(r; E) = \left. \frac{dR_l(r; E)}{dE} \right|_{E=E_l}$ とする. ここで, 式で表される APW 基底関数の球内部分を, ある特定のエネルギー E_l での動径関数とそのエネルギー微分関数により表現する.

$$\phi_n^{\mathbf{k}}(\mathbf{r}) = \frac{1}{\Omega} \sum_{l=0}^{\infty} \sum_{m=-l}^l \phi_{lmn}(|\mathbf{r} - \mathbf{R}|) Y_{lm}(\widehat{\mathbf{r} - \mathbf{R}}) \quad (2.8)$$

$$\phi_{lmn}(\mathbf{r}) = A_{lmn} R_l(r; E) + B_{lmn} \dot{R}_l(r; E) \quad (2.9)$$

\mathbf{K}_n は平面波を区別する逆格子ベクトルを示す. この LAPW 基底関数は, APW 基底関数とは異なりエネルギー依存性を含んでいないことに注意しなければならない. (2.9)式に現れた係数 A_{lmn} と B_{lmn} は, LAPW 基底関数の式の球表面における関数値とその同型微分値が連続となる条件から決定され, 平面波基底の恒等式

$$e^{i\mathbf{k} \cdot \mathbf{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^l i^{-l} j_l(kr) Y_{lm}^*(\hat{\mathbf{k}}) Y_{lm}(\hat{\mathbf{r}}) \quad (2.10)$$

を用いると

$$\begin{aligned}
A_{lmn} &= 4\pi S^2 a_{ln} Y_{lm}^*(\hat{\mathbf{k}}) \\
B_{lmn} &= 4\pi S^2 b_{ln} Y_{lm}^*(\hat{\mathbf{k}}) \\
a_{ln} &= k_n j_l'(k_n S) \dot{R}_l(S; E) - j_l(k_n S) \dot{R}_l(S; E) \\
b_{ln} &= k_n j_l'(k_n S) R_l(S; E) - j_l(k_n S) R_l'(S; E)
\end{aligned} \tag{2.11}$$

ここで、 $\mathbf{k}_n = \mathbf{k} + \mathbf{K}_n$ と略記した。以上から、球内 APW 基底関数（動径関数とそのエネルギー微分関数を用いる）と MT 球外の基底関数（平面波）を用いて、LAPW 基底関数を構築する。

$$\phi_n(\mathbf{r}) = \phi_{lmn}(\mathbf{r}) + \Theta(S - r) \sum_{l=0}^{l_{\max}} \sum_{m=-l}^l [\phi_{lmn}(\mathbf{r}) - \phi_{lmn}^-(\mathbf{r})] \tag{2.12}$$

ここで $\Theta(S - r)$ はヘヴィサイドステップ関数であり、

$$\Theta(x) = \begin{cases} 1, (x > 0) \\ 0.5, (x = 0) \\ 0, (x < 0) \end{cases} \tag{2.13}$$

となる。これにより、エネルギーに依存しない一階微分まで連続な LAPW 基底関数が定義された。

マフィンティン近似を仮定した APW 法や LAPW 法では、図 2.1 のようにポテンシャルの形状を球対称な部分と一定の部分（格子間領域）に分けて計算をしていた。これは、ポテンシャルの形状に制限を加えていることになり、問題がある。こうした制限を取り除き、一般の形状のポテンシャルを計算する手法がある。さらに、価電子だけでなく内殻の電子も考慮し、価電子および内殻の電子がそれぞれ直交化する条件を導入することで、全電子の状態を計算することが可能になった。この手法を FLAPW 法[49, 50]と呼ぶ。

また結晶のような周期的に電荷が分布する系において、Poisson 方程式から静電ポテンシャルを定義する擬電荷胞（Pseudo-charge method）が M. Weinert により提案された[54]。以下にその概要を示す。まず、マフィンティン球内 S における局在電荷分布 $n(\mathbf{r})$ を考える。MT 球外のある点におけるポテンシャル $V(\mathbf{r})$ は、多重極展開[55]により与えられ

$$V(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{4\pi}{2l+1} q_{lm} \frac{Y_{lm}(\hat{r})}{r^{l+1}} \tag{2.14}$$

と表される。ここで q_{lm} は多重極モーメントであり、

$$q_{lm} = \int_S Y_{lm}^*(\hat{r}) r^l \rho(\mathbf{r}) d^3r \tag{2.15}$$

と与えられる。

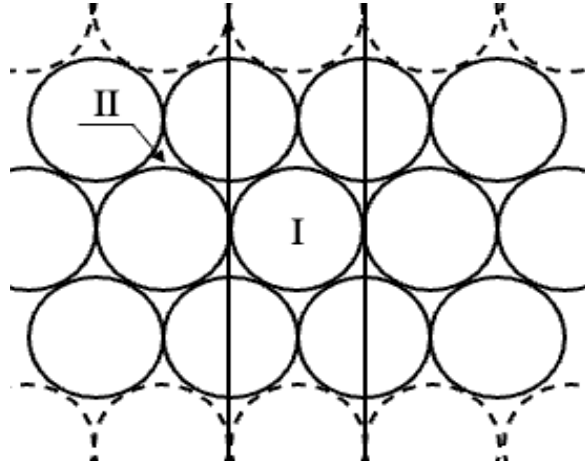


図 2.1 ユニットセルの MT 球 (I) と格子間領域 (II) への分割. Film のような 2 次元系の場合, 点線の球体は真空中に置き換えられる.

ここで正しい多重極モーメントを持つ無限の数の電荷分布の中のいずれかから同じポテンシャル(MT 球外)を得ることができる. 周期系において, 電荷分布は MT 球内および球外の領域にそれぞれ分割することができ,

$$\rho(\mathbf{r}) = \rho_l(\mathbf{r})\theta(\mathbf{r}l) + \sum_i \rho_i(\mathbf{r})\theta(\mathbf{r}_iS_i) \quad (2.16)$$

$\rho_l(\mathbf{r})$ ($\rho_i(\mathbf{r})$)は格子間領域 I (i 番目 MT 球内) における電荷密度を表しており, $\theta(\mathbf{r}l)$ ($\theta(\mathbf{r}_iS_i)$)はそれぞれの領域を分割するステップ関数である. 以下の手順に従って Poisson 方程式から静電ポテンシャルを得る.

- (1) 格子間領域のポテンシャル $V(\mathbf{r})$ を決定する.
- (2) Dirichlet の境界値問題により格子間領域と MT 球との境界を接続する.

はじめに, 格子間領域におけるポテンシャルは, 格子間領域の電荷密度と多重極モーメントのみに依存することから, 電荷密度分布 $\rho(\mathbf{r})$ を MT 球内の擬電荷密度分布 $\tilde{\rho}(\mathbf{r})$ (pseudo charge density) を用いて次式に置き換えることができ,

$$\rho(\mathbf{r}) \rightarrow \tilde{\rho}(\mathbf{r}) = \rho_l(\mathbf{r})\theta(\mathbf{r} \in I) + \sum_i \tilde{\rho}_i(\mathbf{r})\theta(\mathbf{r} \in S_i) \quad (2.17)$$

この擬電荷密度分布は正しい格子間領域のポテンシャルを与えるが, MT 球内においては正しくない. ここで, 式(2.17)を急速に収束するフーリエ級数として表現できると仮定すると

$$\tilde{\rho}(\mathbf{r}) = \sum_{\mathbf{K}} \tilde{\rho}(\mathbf{K}) e^{i\mathbf{K} \cdot \mathbf{r}} \quad (2.18)$$

このフーリエ級数展開から、格子間領域のポテンシャル $V_l(\mathbf{r})$ は、球境界上においてもまた正しいため、MT 球内のポテンシャルは境界値問題となる。次に、Dirichlet の境界値問題から、MT 球内のポテンシャルは Green 関数 G を用いて

$$V_l(\mathbf{r}_i) = \int_{S_i} \rho_i(\mathbf{r}') G(\mathbf{r}_i, \mathbf{r}') d^3\mathbf{r}' - \frac{R_i^2}{4\pi} \oint_{S_i} V_l(\mathbf{R}_i') \frac{\partial G}{\partial n'} d\Omega' \quad (2.19)$$

が得られる。ここで \mathbf{R}_i は MT 球上のある点であり、グリーン関数を用いて、

$$G(\mathbf{r}_i, \mathbf{r}') = 4\pi \sum_{lm} \frac{Y_{lm}^*(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}}_i)}{2l+1} \frac{r_{<}^l}{r_{>}^{l+1}} \left[1 - \left(\frac{r_{>}}{R_i} \right)^{2l+1} \right] \quad (2.20)$$

$$\frac{\partial G}{\partial n'} = \frac{\partial G}{\partial r'} \Big|_{r'=R_i} = -\frac{4\pi}{R_i^2} \sum_{lm} \left(\frac{r_i}{R_i} \right)^l Y_{lm}^*(\hat{\mathbf{r}}') Y_{lm}(\hat{\mathbf{r}}) \quad (2.21)$$

である。ここで、 $r_{>}$ ($r_{<}$) は、 r_i と r の大きい方 (小さい方) であり、正規微分であらわされる。ただし、(2.19)式中において実際の電荷密度を用いていることから、どこでも正しいポテンシャルを持つ。(2.20)式から(2.21)式は、Poisson 方程式の正確な解を与えており、擬電荷密度分布のフーリエ展開の収束が早いことから、MT 球内および球外のそれぞれの領域に分割した本手法において有効的な計算手法と言える。

2.3 +U 法による電子強相関効果補正項

LDA や GGA を用いたバンド理論は、これまでに s , p 軌道上の電子を中心として、金属や半導体材料の電子構造の理解に大きな成功を収めてきた。しかし d , f 軌道上の局在した電子を含む強相関電子系においては未だ不完全である。特に $3d$, $4f$ 軌道を占有する電子は、配位子との間で軌道の重なりが小さくなることにより、原子核付近への局在性が強まり、遮蔽効果が不完全となる。これにより LDA や GGA といった平均場近似を前提としたバンド理論は正しい結果を与えなくなる。この問題を改善するために提案されたのが、電子間の相互作用を考慮する Hubbard 模型[56-61]に基づいた DFT+U 法[3]である。Hubbard 模型では、ある特定の原子上の電子間の反発力は、Hubbard のオンサイトクーロン相互作用パラメータにより定量化される。さらに Hubbard 模型に基づき、非磁性系における不純物の磁気特性について研究するために提案された Anderson 模型[62, 63]の成功に触発され、Anisimov ら[64]が通常の DFT のエネルギー汎関数に対して、軌道に依存性を持つ追加のエネルギー補正項を提案している。これが DFT+U 法であり、具体的には局在した $3d$, $4f$ 軌道上の電子に対して、遮蔽された Hartree-Fock 型の電子間の相互作用が導入される。

DFT+U 法における全エネルギー[63]は次式で表される。

$$E_{\text{tot}}[\rho(r), \hat{\rho}] = E_{\text{DFT}}[\rho(r)] + E_{ee}(\hat{\rho}) - E_{dc}(\hat{\rho}) \quad (2.22)$$

(2.22)式において、第一項 $E_{\text{DFT}}[\rho(r)]$ は全電子スピン密度 $\rho^\sigma(r)$ ($\sigma = \uparrow, \downarrow$) の全エネルギー汎関数である。第二項 $E_{ee}(\hat{\rho})$ は、軌道に依存する電子間相互作用エネルギー、第三項 $E_{dc}(\hat{\rho})$ は、すでに第一項 $E_{\text{DFT}}[\rho(r)]$ において含まれている電子間相互作用エネルギーを取り除く double-counting 項である。これらは、原子極限の軌道 $\{\phi_m\}$ (m は磁気量子数) における局所軌道占有行列 $\hat{\rho}^\sigma = \hat{\rho}_{mm'}^\sigma$ の汎関数となっている。電子間相互作用エネルギー $E_{ee}(\hat{\rho})$ は

$$E_{ee}(\hat{\rho}) = \frac{1}{2} \sum_{m_1, m_2, m_3, m_4} \sum_{\sigma, \sigma'} \rho_{m_1, m_2}^\sigma [\langle m_1, m_3 | V_{ee} | m_2, m_4 \rangle - \langle m_1, m_3 | V_{ee} | m_2, m_4 \rangle \delta_{\sigma, \sigma'}] \rho_{m_3, m_4}^{\sigma'} \quad (2.23)$$

である。 V_{ee} は有効オンサイトクーロン相互作用、 $\langle \rangle$ は角度積分を意味している。また、 ρ^σ はオンサイトにおける $d(f)$ 軌道の電子占有行列である。原子極限における電子間相互作用ポテンシャルは、 $d(f)$ 軌道の球面調和関数を用いて、

$$\langle m_1, m_3 | V_{ee} | m_2, m_4 \rangle = \sum_k a_k(m_1, m_2, m_3, m_4) F^k \quad (2.24)$$

$$a_k(m_1, m_2, m_3, m_4) = \frac{4\pi}{2k+1} \sum_{q=-k}^k \langle Y_{lm_1} | Y_{kq} | Y_{lm_2} \rangle \langle Y_{lm_3} | Y_{kq}^* | Y_{lm_4} \rangle \quad (2.25)$$

と表される。ただし、 $0 \leq k \leq 2l$ である。ここでクレブシュ-ゴルダン係数を用いて次式のように表される。

$$\langle Y_{lm_1} | Y_{kq} | Y_{lm_2} \rangle = (-1)^m \left[\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi} \right]^{\frac{1}{2}}$$

$$\times \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix} \quad (2.26)$$

また式(2.24)においてスレーター積分 F^k は一般的に次式のように定義される.

$$F^k = \int_0^\infty r_1^2 r_2^2 R_{nl}^2(r_1) R_{nl}^2(r_2) \frac{r_{<}^k}{r_{>}^{k+1}} dr_1 dr_2 \quad (2.27)$$

ここで $r_{<}$ 及び $r_{>}$ は r_1 と r_2 のうち,それぞれ大きい方と小さい方を意味しており, $R_{nl}(r)$ は主量子数 n , 軌道角運動量量子数 l の時の動径関数である. (2.26)式から, $m_1 = m_2 = m_3 = m_4$ かつ $\sigma = \sigma'$ のとき, 第一項における自己相互作用が第二項の交換相互作用により完全に相殺されていることが分かる. また double-counting 項は次式のように表される.

$$E_{dc}(\hat{\rho}) = \frac{U}{2} \rho(\rho - 1) - \frac{J}{2} \rho^\sigma(\rho^\sigma - 1) \quad (2.28)$$

ここでオンサイトクーロン相互作用と交換相互作用 J はそれぞれ次式のように与えられる.

$$U = \frac{1}{(2l+1)^2} \sum_{m_1, m_3} \langle m_1, m_3 | V_{ee} | m_1, m_3 \rangle \quad (2.29)$$

$$J = U - \frac{1}{2l(2l+1)} \sum_{m_1, m_3} \langle m_1, m_3 | V_{ee} | m_1, m_3 \rangle - \langle m_1, m_3 | V_{ee} | m_3, m_1 \rangle \quad (2.30)$$

これらは, d 電子に関しては F^0 , F^2 および F^4 のスレーター積分で表される. また $n^\sigma = \text{Tr} \hat{n}^\sigma$, $n = \sum_\sigma n^\sigma$ は全てのオンサイトにおける $d(f)$ 占有電子数である. 本論文では簡単のために $J=0$ として, J の効果を取り込んだ有効オンサイトクーロン相互作用パラメータ $U_{\text{eff}} = U - J$ として再定義している.

2.4 +U 法の物理的解釈

初めに、前節で説明した局在軌道における電子間に働く相互作用は、交換相互作用 J よりもクーロン相互作用 U がはるかに大きい。したがって、以降の議論では式(2.30)において $J=0$ を仮定し、電子間に働く球対称な有効オンサイトクーロン相互作用 U_{eff} のみに注目する。よって+U法での全エネルギーは式(2.22)および式(2.28)より

$$\begin{aligned} E^{+U}(\hat{\rho}) &= E_{ee}(\hat{\rho}) - E_{dc}(\hat{\rho}) \\ &= \frac{U}{2} \sum_{m,\sigma} \rho_{mm}^{\sigma} - \sum_{m,\sigma} \rho_{mm'}^{\sigma} \rho_{m'm}^{\sigma} \\ &= \frac{U}{2} \text{Tr}[\rho^{\sigma}(1 - \rho^{\sigma})] \end{aligned} \quad (2.31)$$

と書き換えられる。電子数が変動し続ける開放系において、系の波動関数は純粋な状態の波動関数ではなく、複数の状態が統計的に混合されたものとして記述される。例えば、 $N + \omega$ 個の電子をもつ系の全エネルギーは(N は整数で、 $0 \leq \omega \leq 1$ とする)

$$E_n = (1 - \omega)E_N + \omega E_{N+1} \quad (2.32)$$

で表される。ここで E_N と E_{N+1} は、それぞれ N 個と $N+1$ 個の電子をもつ状態の系のエネルギーを示しており、 ω は、 $N+1$ 個の電子をもつ状態の統計的な重みを表している。したがって開放系における全エネルギーは、図 2.1 に示すようにある原子における局在軌道の占有電子が整数となる状態を一連の直線で結んだ形となる。その代わりに、エネルギーと占有電子数で表される曲線の勾配は、整数個の電子の不連続性を伴う区分的に一定なものであり、 $N(N+1)$ 個の電子系の電子親和力(イオン化ポテンシャル)に対応する。

一方で、平均場近似を仮定した LDA(GGA)法に基づいた DFT 計算では、このような区分的に直線な値は得られず、系の局在軌道に対する占有電子数が非整数をとる際に、全エネルギーは、占有電子数に対する非物理的な曲率とともに誤った最小値を与えてしまう。そのような非物理的な曲率は、占有軌道の全エネルギーに対して非線形な寄与を与える部分的に占有された Kohn-Sham 軌道の自己相互作用による LDA(GGA)法による誤った取り扱い(主に交換相関項における適切に相殺されない Hartree エネルギーから生じる 2 次項)と関連している。しかしながら、軌道の占有電子数が整数値であると仮定すること、すなわち拘束されることにより、LDA(GGA)法を用いて極めて正確に得られることが知られている[65]。また、占有電子数が非整数をとるすべての範囲において LDA(GGA)法の全エネルギーの概算の曲率(図の”LDA+U correction”)を除去するような補正を加えることで、物理的状态(ほぼ区分的な線形の総エネルギー曲線)を得られることを意味している。したがって、クーロン相互作用パラメータ U は、占有電子数 N を関数とした LDA(GGA)法の全エネルギーの非物理的な曲率として考えることができ、系に誤って計上された非整数個の占有数を持つ電子の自己相互作用に関連している。このような状況は、凝集系においてより複雑化することが考えられる。結晶場(配位子場)下において、局在した原子軌道間で混成軌道が形成されて

しまうため, この軌道の混成効果も LDA(GGA)法の全エネルギーから取り除かなければならない.

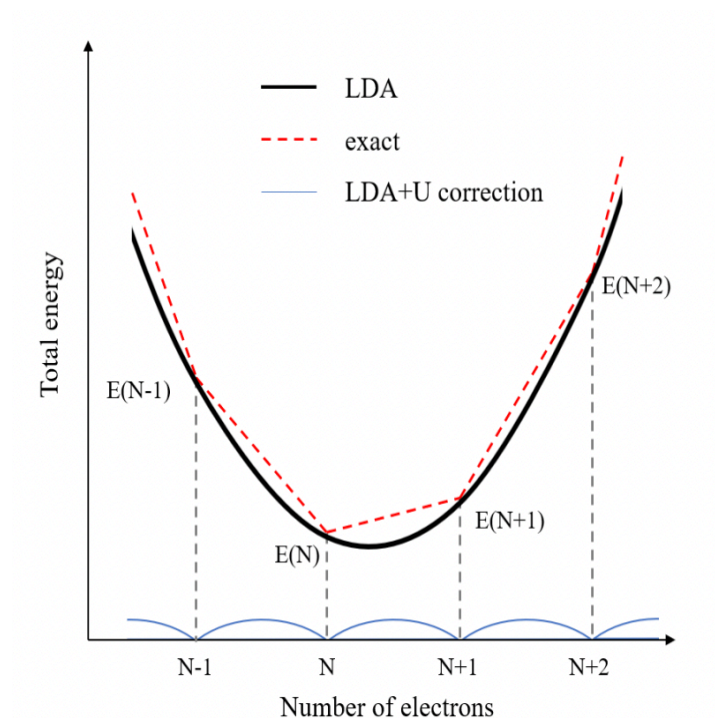


図 2.1 一般的な解放系における電子が局在したあるサイトでの電子数に対する全エネルギーの関数黒の実線および赤の波線は, それぞれ, LDA 法により得られる非物理的な全エネルギーの曲線と厳密な全エネルギーを示している. これらの差分を青の実線で示しており,これが+U法によって補正されるべき部分である.

2.5 第一原理計算に基づく誘電関数と磁気光学カー効果の計算

誘電率は物質の光学応答を解析する場合に用いられ、一般に両者の間には簡単な関係が成り立つ。しかし一般には金属を扱う場合は伝導率 σ を、絶縁体であれば誘電率 ϵ を用いる。これは金属の誘電率 ϵ が光のエネルギー ω が $\omega \rightarrow 0$ の極限（静電場）において自由電子の遮蔽効果によって発散してしまうのに対し、伝導率 σ は有限の値に収束するため、都合が良いからである。

誘電率は電気分極 P と電場 E の関係を与える量であり、両者の関係は次式で表される。

$$P = \hat{\epsilon} E \quad (2.33)$$

$$P_i = \epsilon_{ij} E_j \quad (2.34)$$

ここで誘電率テンソル ϵ は

$$\hat{\epsilon} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{pmatrix} \quad (2.35)$$

で表される。一般に誘電率テンソルの成分は複素数であるため、

$$\epsilon_{ij} = \epsilon_{ij}^{(1)} + i\epsilon_{ij}^{(2)} \quad (2.36)$$

と表すことにする。

また ϵ の実部と虚部の間にはクラマース・クローニッヒの関係

$$\epsilon_{ij}^{(1)} = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\epsilon_{ij}^{(2)}(\omega')}{\omega' - \omega} d\omega' \quad (2.37)$$

$$\epsilon_{ij}^{(2)} = -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\epsilon_{ij}^{(1)}(\omega')}{\omega' - \omega} d\omega' \quad (2.38)$$

が成り立つ(\mathcal{P} はコーシー主値)ため、実部・虚部のうち一方のスペクトルが測定できれば、もう一方のスペクトルも計算することができる。

磁気光学カー効果(Magneto-optical Kerr Effect; MOKE)は直線偏光した光が磁性体表面で反射した時、反射光が楕円偏光へと変化し、主軸が回転する現象である。その元々の偏光方向に対する反射光の回転角をカー回転角と呼ぶ。この角度は一般的に表面の形状や磁化の方向、入射光の波数ベクトルの方向等に依存する。また磁気光学効果は磁化の方向により極磁気カー効果（磁化が反射面の法線方向）、縦カー効果（磁化が試料面内かつ入射面内）、横カー効果（磁化が試料面内）に分類されるが、本研究では物理的関心が高い極磁気カー効果について議論する。この場合複素カー回転角は巨視的な伝導率の観点から表されることが知られており[66]、 z 軸より入射する光に対する立方晶系における複素磁気光学カー効果は次式で与えられる[67]。

$$\Phi_K \equiv \phi_K + i\epsilon_K = \frac{-\sigma_{xy}}{\sigma_{xx}\sqrt{1 + i\left(\frac{4\pi}{\omega}\right)\sigma_{xx}}} \quad (2.39)$$

ここで、 ϕ_K はカー回転角と呼ばれる実部であり、 ϵ_K はカー楕円率と呼ばれる虚数部である。また σ_{xx} と σ_{xy} はそれぞれ光学伝導度テンソルの対角成分と非対角成分である。また巨視的な光学伝導度テンソルは久保公式によって微視的な光学遷移と関連づけられることが知られている[68]。ここで式におけるバンド間遷移の σ_{xx} と σ_{xy} の成分の計算については、Oppeneier らによって提案された計算手法[69]を採用している。以下に簡単に示す。

$$\sigma_{xy}(\omega) = \frac{ie^2}{m^2\hbar} \sum_{\mathbf{k}} \sum_{\ell\sigma_1} \sum_{\substack{n\sigma_2 \\ \text{OCC UNOCC}}} \frac{1}{\omega_{n\sigma_2\ell\sigma_1}(\mathbf{k})} \left(\frac{\Pi_{n\sigma_2\ell\sigma_1}^x \Pi_{n\sigma_2\ell\sigma_1}^y}{\omega - \omega_{n\sigma_2\ell\sigma_1}(\mathbf{k}) + \frac{i}{\tau}} + \frac{\left(\Pi_{n\sigma_2\ell\sigma_1}^x \Pi_{n\sigma_2\ell\sigma_1}^y\right)^*}{\omega + \omega_{n\sigma_2\ell\sigma_1}(\mathbf{k}) + \frac{i}{\tau}} \right) \quad (2.40)$$

$$\sigma_{xx}(\omega) = \frac{ie^2}{m^2\hbar} \sum_{\mathbf{k}} \sum_{\ell\sigma_1} \sum_{\substack{n\sigma_2 \\ \text{OCC UNOCC}}} \frac{1}{\omega_{n\sigma_2\ell\sigma_1}(\mathbf{k})} \left(\frac{|\Pi_{n\sigma_2\ell\sigma_1}^x|^2}{\omega - \omega_{n\sigma_2\ell\sigma_1}(\mathbf{k}) + \frac{i}{\tau}} + \frac{|\Pi_{n\sigma_2\ell\sigma_1}^x|^2}{\omega + \omega_{n\sigma_2\ell\sigma_1}(\mathbf{k}) + \frac{i}{\tau}} \right) \quad (2.41)$$

ここで $1/\tau$ はバンド間遷移に伴う電子散乱の緩和時間を意味しており、スペクトルが実験に一致するように適切な値を選択する。(2.40)式と(2.41)式において、ある非占有軌道のバンド n とそのバンドにおけるスピン σ_2 、ある占有軌道におけるバンド ℓ とそのバンドにおけるスピン σ_1 を用いて次のような占有軌道及び非占有軌道のエネルギー差 $\hbar\omega_{n\sigma_2\ell\sigma_1} = E_{n\sigma_2}(\mathbf{k}) - E_{\ell\sigma_1}(\mathbf{k})$ が表されている。すなわち占有軌道と非占有軌道における電子遷移を考慮することで光学伝導度を計算しているといえる。またさらに式及び式における運動量演算子の行列要素 $\Pi_{n\sigma_2\ell\sigma_1}$ は次式で表される。

$$\Pi_{n\sigma_2\ell\sigma_1} = \int \psi_{nk\sigma_2}^*(\mathbf{r}) \left[\mathbf{p} + \left(\frac{\hbar}{4mc^2} \right) [\boldsymbol{\sigma} \times \nabla V(\mathbf{r})] \right] \times \psi_{\ell k\sigma_1}(\mathbf{r}) d\mathbf{r} \quad (2.42)$$

ここで $\psi_{nk\sigma_2}(\mathbf{r})$ はブロッホ波動関数である。またここで式においてスピン軌道相互作用 (spin orbit coupling: SOC) 項は、正準運動量演算子 \mathbf{p} に比べて遥かに小さいため無視できる。これは、スピン反転による遷移を考慮しないことを意味しているがスピン軌道相互作用自体については運動エネルギー項及び波動関数を通じて考慮されている。また伝導率テンソルに対するバンド間の寄与のみならずバンド内の寄与もまた考えられるが、次式で表される経験的なドルーデモデルによって適切に表される。

$$\sigma_D(\omega) = \frac{\sigma_0}{1 - i\omega\tau_D} \quad (2.43)$$

通常、定数である σ_0 と τ_D については実験から得られた値が使われる。

ここまで第一原理計算における標準的な誘電関数と磁気光学カー効果の理論についてまとめた。実際の計算には伝導率テンソルの対角成分 $\sigma_{xx}(\omega)$ と非対角成分 $\sigma_{xy}(\omega)$ を計算する際のブリルアンゾーン積分の問題や運動量演算子の行列要素 $\Pi_{n\sigma_2\ell\sigma_1}$ を計算する際に必要な

規定関数の問題が存在し，これらを解決する手法については前述の Oppeneer らの報告[69] に詳細が記述されている．

2.6 おわりに

本章では、密度汎関数理論の基礎である Hohenberg-Kohn の定理について言及し、定義された全エネルギー汎関数から Kohn-Sham 方程式を導いた。さらに Kohn-Sham 方程式を解くために必要となる交換相関項の近似について、局所スピン密度近似および一般化勾配近似の二つを示した。さらに本研究で取り扱う第一原理バンド計算手法である FLAPW 法について説明した。また本研究で取り扱う強相関電子系に対し、有効な計算手法である DFT+U 法について言及した。最後に Oppeneer の理論に基づく光学応答について、材料の磁性に対して第一原理計算から計算するために必要なスピン軌道相互作用の導入、そして磁気光学効果の計算手法について概説した。

第3章 ニューラルネットワークによる機械学習

3.1 はじめに

ニューラルネットワーク(Neural Network; NN)は動物の脳の動作を数理モデルで模倣したもので、脳細胞をつなぐシナプスの結合状態と電気信号による情報の伝播を数値の演算で再現したものである。NNのもとになる2入力1出力の数理モデルであるパーセプトロンは1958年にRosenblattらが発表した[70]。パーセプトロンは、真が1、偽が0である入力に対し重みと呼ばれる係数でそれぞれ乗算してその結果を総和し、一定の閾値以上の結果であれば1を、閾値以下であれば0を出力するものである。これは線形な論理演算（例えば論理和(OR)、論理積(AND)など）に対して適切な重みを選ぶことで実現することができたが、非線形な論理演算（例えば排他的論理和(XOR)）を実現することは困難であった。その後1980年代にパーセプトロンを層状に構築し、これらの間を活性化関数による非線形変換で結合した多層パーセプトロン[71, 72]が開発された。これにより入力に対し非線形な出力を行うことが可能になり、また論理演算以外に実数演算にも拡張することが可能になった。1986年にはRumelhartらが多層パーセプトロンの学習に必要な勾配計算を効率化する誤差逆伝播法(backpropagation)[73]を発表した。これにより3層パーセプトロンの現代的な学習手法が完成し、2000年代初頭に実用的な非線形予測手法として一定の成功を収めた。しかし時代を経るにつれ世界に存在するデータが爆発的に増加し、従前より複雑かつ巨大データの塊であるビッグデータによる予測の必要性が叫ばれ始めた。しかし予測の難易度が上昇し、また当時のコンピュータの計算速度では演算時間が現実的でなかったため、一時NNを含めた機械学習は下火になっていた。しかし2006年HintonとSalakhutdinovがニューラルネットワークを用いた高次元ベクトルデータの次元削減手法であるAuto Encoder[74]を発表し、データの特徴を維持しつつ低次元空間に圧縮する手法が注目を集め、またGeneral Purpose computing on Graphics Processing Units(GPGPU)による演算高速化が実現されたことから再び機械学習が注目を集めるようになった。これ以降はビッグデータに対応可能なより層の多い深層学習(deep learning)が主流となっている。現在では多層パーセプトロンによる全結合ニューラルネットワークのほか、画像認識・分類で用いられる畳み込みニューラルネットワーク(Convolutional Neural Network; CNN)[75]、自然言語処理で用いられる再帰型ニューラルネットワーク(Recurrent Neural Network; RNN)[73]などがしばしば用いられている。このような発展から数ある機械学習アルゴリズムの中でもニューラルネットワークは重要な一角を占めるようになった。また材料設計や脳神経科学、信号処理などさまざまな分野での応用が進んでいるため今後の発展がますます期待されている。

本章ではニューラルネットワークを用いた訓練・推論方法, 訓練時におけるハイパーパラメータの調整方法とその評価方法を記述する.ニューラルネットワークで構築されたモデルの訓練は

- ①データの前処理
- ②訓練データによるモデルの訓練とハイパーパラメータの調整
- ③テストデータによる汎化性能の評価

の順で実施される. まず①では, 訓練・テストにおけるデータを機械学習に適した形に変換する方法を述べる. また前処理の一環としてデータの少なさを補うデータ拡張についても記述する. 次に②では, モデルの重み行列を, 誤差逆伝播法と重み更新アルゴリズムに基づき最適化する方法について述べる. また訓練においてはハイパーパラメータと呼ばれるパラメータがモデルの性能に大きく影響しており, これらのチューニング方法と評価指標についても記述する. 最後に③では, 訓練時には用いなかったデータを用いてモデルの汎化性能を評価する方法を述べる.

3.2 データの前処理と訓練・テストデータの分割

NNに限らず機械学習に用いるベクトルのデータセットは、生データのままでは学習に適さないことが多い。例えば表形式のデータであればデータ欠損が存在する場合や、波形、画像などであればデータごとに値の範囲が異なる、一つのデータのベクトルに対し次元が高すぎることから機械学習の推論において重要な特徴量を取り出せない場合がある。そのためモデルの学習に適した形式に変換するデータの前処理が必要である。表に主なデータの前処理手法を挙げる。

データの前処理は①欠損値に対する処理②データスケーリング③次元削減の順にて行われる。

まず欠損値に対する処理は、欠損値の部分の活用・無視を判断することから始まる。機械学習の訓練・テストに用いるデータ数が十分に存在し、データに偏りが無い場合は欠損値を含むデータは全て除外して学習を行う。しかしデータ数が少数である場合や偏りが存在する場合は欠損値を何らかの手法で予測し、穴埋めする必要がある。予測手法には周辺のデータの傾向から内挿する単一代入法や多重代入法、統計的手法である完全情報最尤法などの手法で値を予測する。

データスケーリングは機械学習に用いる特徴量の値の範囲を一定の範囲に揃える手法である。これは異なる範囲の特徴量を入力すると大きな値の範囲を持つ特徴量がモデルの学習において支配的になり他の特徴量による寄与が無視されてしまうためである。主な手法としては正規化 (standardization, min-max scaling) と標準化 (normalization, z-score scaling) が挙げられる。正規化はデータの範囲を 0 から 1 へ変換する手法であり、次式で表される。

$$\hat{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3.1)$$

ここで $x = \{x_0, x_1, \dots, x_N\}$ は正規化前の特徴量の各値、 \min , \max はそれぞれ x 中の最小値、最大値を返す関数である。この手法は主に特徴量のデータ分布がすべての特徴量に対し同じ場合に行われることが多い。次に、標準化は全ての特徴量の平均 μ を 0、分散 σ^2 を 1 に変換する手法であり、次式で表される。

$$\hat{x} = \frac{x - \mu(x)}{\sigma^2(x)} \quad (3.2)$$

ここで μ, σ^2 はそれぞれ x における平均と分散であり、次式にて計算される。

$$\mu = \frac{1}{N} \sum x_i \quad (3.3)$$

$$\sigma^2 = \frac{1}{N} \sum (x_i - \mu)^2 \quad (3.4)$$

最後に次元削減は特徴量の次元をアルゴリズムに沿って削減する手法である。一般に機械学習モデルのフィッティングは入力の特徴量の次元が大きくなるほど性能が悪くなる

「次元の呪い」という現象が発生する。これは特徴量の次元が大きくなるほど機械学習内部の変数も多くなり複雑な推論ができるようになる一方、予測には寄与の小さい特徴量も学習してしまうため過学習を起こしてしまう現象である。そのため特徴量を機械学習の予測に対する寄与が大きいものを抽出し入力ベクトル次元を小さくする手段が取られる。主な手法としては主成分分析(Principal Component Analysis; PCA)[76]や Auto Encoder が用いられる。PCA は特徴量のベクトルから共分散行列を生成し、その固有値から共分散行列の特徴づけに対して寄与の大きい固有ベクトルを数個抽出し新たな特徴量のベクトルの各要素の軸として用いる手法である。また Auto Encoder は中央の隠れ層の次元を小さくした砂時計型 NN を用いた教師なし学習であり、入力データを中間層で低次元ベクトル空間へ圧縮し、そこから高次元の出力層へデータを再変換する。これにより高次元データを低次元ベクトル空間へ可逆圧縮する変換器を作ることができ、NN により構成されることから PCA と比較して重みの直交性の制約が緩い変換を行うことができる。

本研究では①と③は、データに欠損値がないこと、次元削減することが難しいことから行わず、②のデータスケーリングのみ行なった。また活性化関数に ReLU 関数（詳細は後述）を用いたことから負値を入力することができないため、正規化のみ実施した。

機械学習では、元のデータを一定の割合で分割しモデルの訓練に用いる訓練データとモデルの汎化性能を測るためのテストデータに分割する。分割の比率は訓練データ：テストデータ=7：3 もしくは 8：2 がよく用いられており、訓練データにデータ数の比重を多くする。また訓練データは後述するミニバッチ学習に用いるためデータの順番をシャッフルする必要がある。

3.3 データ拡張による予測モデル性能向上

機械学習の予測精度は、モデルの構成とハイパーパラメータの調整のみならず、データ数とデータの品質にも左右される。そのため予測するデータのサンプリングやラベルに偏りがなく、可能な限り多くデータを集める必要がある。例えば、画像分類問題にてモデルの評価に頻繁に用いられるデータセットである MNIST[77]では 0~9 までの手書き文字を 7 万データ収録している。他にもこの世に実在するモノ（飛行機や自動車など）の画像を収録した CIFAR10[78]では約 6 万枚ものデータを収録している。しかしながら機械学習を行うすべての場合において膨大なデータが得られるとは限らず、限られた数でデータセットを組む場合がある。例えば画像認識であれば対象のデータの収集や、機械学習に読み込ませるためのトリミングや回転などの前処理などは人力で行う必要があり、膨大な数のデータに対してそれを行うことは極めて高コストである場合が多い。そのため必然的にデータ数は MNIST や CIFAR10 と比べて不足することになり、このことが原因で予測モデルの精度が出ない場合がある。そこで、少数データでの予測モデルの高精度化のため、既存のデータをもとに新しいデータを作成し訓練データを水増しするデータ拡張法(Data Augmentation; DA)[79]を行う。データ拡張は主に分類問題の画像認識で広く使われており[80-82]、例えばオリジナルの画像に対して拡張したデータを加えることで分類正答率を向上させる試みが報告されている。この場合、DA には画像の回転、縮尺、切り抜きなどが報告されており、性能の向上が報告されている。

一方回帰問題に対してはデータが 1 次元であることが多いため、回転など画像に対する処理方法を適用することが難しい。そこでガウス分布に基づくノイズを付加することによりオリジナルのデータに似通った手法を用いる DA が採用されている。先行研究での報告は多くなく、時系列データを取り扱う LSTM モデルにおいて性能が 10%程度向上したという報告[83]がなされており、現在注目を集めつつある。

本研究で用いるガウスノイズによる DA は次式のように訓練データの入力のみに対してかける。

$$\mathbf{x}_{\text{DA}}^{\text{train}} = \mathbf{x}_{\text{original}}^{\text{train}} + \lambda \epsilon_i \quad (3.5)$$

$$\mathbf{x}^{\text{train}} = \mathbf{x}_{\text{original}}^{\text{train}} + \mathbf{x}_{\text{DA}}^{\text{train}}_0 + \dots + \mathbf{x}_{\text{DA}}^{\text{train}}_N \quad (3.6)$$

ノイズ ϵ は発生確率が正規分布 $N(0,1)$ に従う乱数である。またスケール変数 λ はガウス分布の確率分散に相当し、乱数の大きさを決めるハイパーパラメータである。本研究ではこの他にも学習回数、拡張データ倍率もハイパーパラメータとして取り扱う。またノイズは乱数を変えていくつか生成しそれぞれの拡張後データにおいて完全には一致しないように生成する。

3.4 ニューラルネットワークの訓練方法

全結合型ニューラルネットワークは図 3.1 のモデルのように特徴量入力ベクトル, 出力をもつ非線形ブラックボックス関数である. 図 3.1 中の○の部分にはノードと呼び, 各変換段階において入力ベクトルを入力する部分を入力層, 出力ベクトルを出力する層を出力層, その中間の層を隠れ層という.

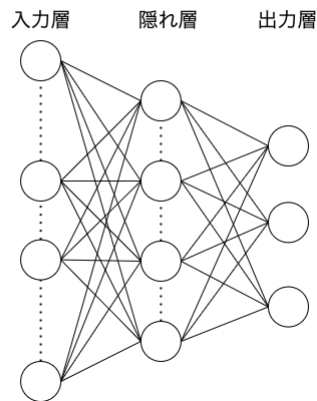


図 3.1 3 層ニューラルネットワークの模式図. 各○がノード, 各線がノード間の接続を表す. 計算の流れは左から右に進行する.

その中身は各層間に構成した重み行列・バイアスベクトルによる **Affine** 変換と非線形活性化による変換であり, 適切な重み行列・バイアスベクトルを構成することで目的の入出力関係を再現することができる. 図 3.1 のモデルを数式で表すと次式となる.

$$\mathbf{h} = f(W_0 \mathbf{x} + \mathbf{b}_0) \quad (3.7)$$

$$\mathbf{y} = f(W_1 \mathbf{h} + \mathbf{b}_1) \quad (3.8)$$

W_i は重み行列, \mathbf{b}_i はバイアスベクトル, \mathbf{h} は隠れ層のベクトル, $f(\cdot)$ は活性化関数である. 活性化関数にはさまざまな式が提案されており, 例えば **tanh** 関数, **sigmoid** 関数などが挙げられる. しかしこれら両者は入力が-1 から 1 の間で変化することを期待した関数であり, それ以外では一定値に収束するという性質を持っている. そのため後述する誤差逆伝播法による勾配計算の際, -1 から 1 の範囲を外れた場合, その勾配はほとんど 0 になるという性質を持っており, もし勾配が 0 であるならばそれよりも前の層の勾配も全て 0 になり, 重み更新がされない勾配消失問題が発生する. そのため 2010 年, Nair らが次式のような正規化線形関数(**Rectified Linear Unit**; **ReLU**)[84]を発表し, 今日では多くの予測モデルで活性化関数として広く使われている.

$$\text{ReLU}(x) = \max(0, x) \quad (3.9)$$

$\max(0, x)$ は入力 x に対し $x > 0$ ならば x , $x \leq 0$ ならば 0 を返す関数である. 図 3.2 に入力に対する応答を示す.

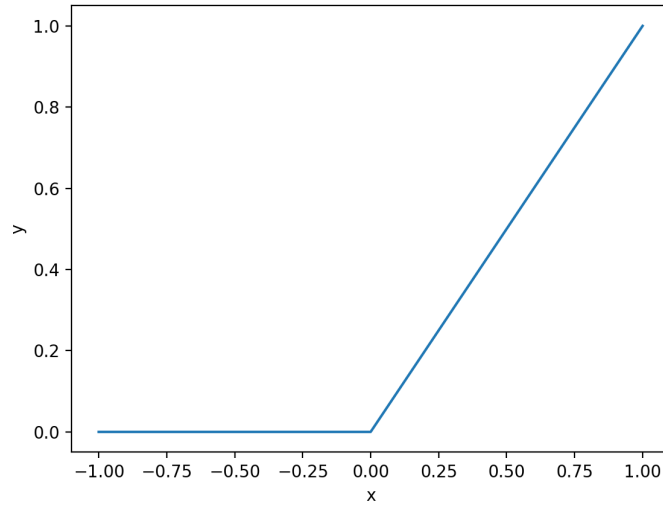


図 3.2 ReLU 関数の入出力関係. 横軸が入力, 縦軸が ReLU 関数の出力を示す.

ReLU の微分は $x > 0$ ならば 1, $x \leq 0$ ならば 0 となり, 入力为正の値であるときのみ勾配を持つ. これにより入力の値が大きすぎるために発生する勾配消失問題を解決することができ, かつ負の入力を遮断する非線形を保っている.

NN における「訓練」とは, 訓練データを用いて入力から出力までの非線形変換に用いるモデルの重み行列をアルゴリズムに沿って最適化することである. すなわち訓練データ $\mathbf{x}^{\text{train}}$, $\mathbf{y}^{\text{train}}$ 用いて以下のようなモデルによる予測値と正解の出力値の誤差を最小化する W を求めることである.

$$\operatorname{argmin}_W \left(L(\mathbf{y}^{\text{train}}(W, \mathbf{x}^{\text{train}}), \mathbf{y}^{\text{pred}}) \right) \quad (3.10)$$

L は損失関数(loss function)と呼ばれ, 回帰問題では平均二乗誤差(Mean squared error; MSE)

$$L(\mathbf{y}^{\text{train}}, \mathbf{y}^{\text{pred}}) = \sum_i^N \frac{1}{N} (\mathbf{y}^{\text{train}} - \mathbf{y}^{\text{pred}})^2 \quad (3.11)$$

によって計算される.

前処理済み訓練データと初期化されたモデルは準備済みと仮定した NN の訓練段階におけるアルゴリズムは以下の通りである. 計算フローも図 3.2 に示した.

1. 重み更新のため, 訓練データをミニバッチに分割する.
2. 全てのミニバッチを用いてモデルの訓練を行う.
 - 2.1 1 つのミニバッチの入力データを取り出してモデルに入力し, モデルの出力値を得る.
 - 2.2 モデルの出力値と正解データを損失関数へ入力し誤差を計算する.
 - 2.3 誤差とモデルの重みから重みに対する損失の勾配 $\nabla_W L$ を計算する.
 - 2.4 勾配と重み更新式に基づき, W を更新する.

2.5 指定のエポック数まで 1-2 を繰り返す.

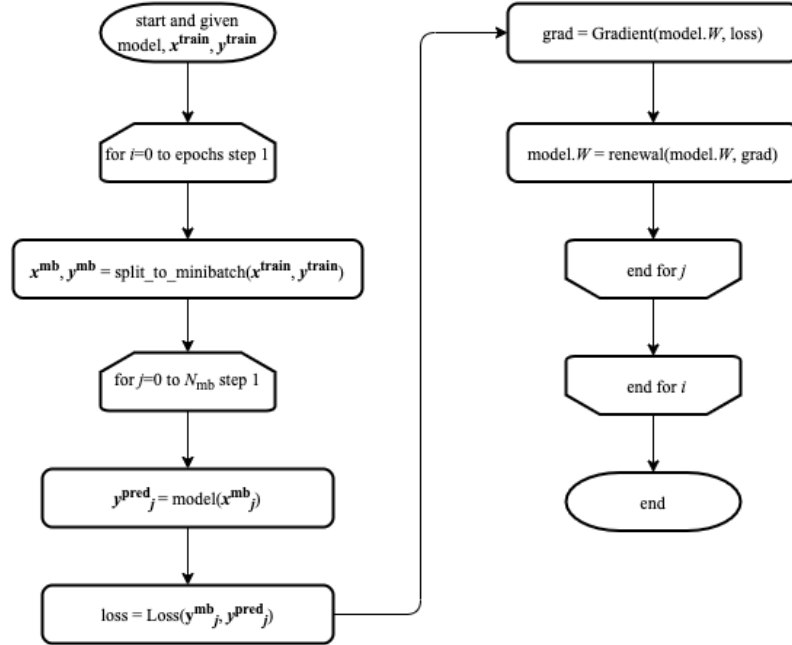


図 3.2 訓練過程の計算フロー. 可視化のためミニバッチへの分割や損失関数など特殊処理に関しては関数で表現した. また `model` は重み `model.W` の要素を持つインスタンスである.

なお, 実装において訓練データ・テストデータの分割する時の組み合わせやミニバッチを生成する時の組み合わせ, モデルの重みの初期値は疑似乱数を用いてランダムに決定される. 本研究で用いた疑似乱数はメルセンツツイスタ法[85]によって生成されており, 計算データの再現性を担保するために乱数の初期値 (乱数シード) を固定する.

3.4.1 重み更新式

NN の重みを最適化は, 複雑な重みの探索空間にて式(3.10)を満たすローカルミニマムを探索することである. 従来 W の更新には, η を学習率, L を損失関数とした最急降下法

$$W_t = W_{t-1} - \eta \nabla_W L \quad (3.12)$$

が用いられてきた. この手法は t 回目の重み W を, $t-1$ 回目の重み W_{t-1} に対し一定値の学習率分だけ損失関数の勾配 $\nabla_W L$ を用いて重みを変化させるものである. 学習率はハイパーパラメータであり, 人為的に決定される. 最急降下法のデメリットは L から計算される勾配を全て用いて最適化しているため, 一度最適でないローカルミニマムに陥ると脱出は困難なことである. そこで重み更新のたびに訓練データをランダムに1つ選択し, 毎回違うデータで勾配計算を行うことで最適でないローカルミニマムに陥る可能性を抑えた確率的勾配降下法(Stochastic Gradient Descent, SGD)[86]が開発された. しかし SGD は 1 データごと

に重みを更新する非効率な手法であるため、ミニバッチと呼ぶランダムに選択したデータの集合を用いて計算を行う。ミニバッチ学習における重み更新は 1 ミニバッチ内全ての訓練データを用いて重み更新を行うことで高速化をはかる手法であり、SGD に必要な重み更新に用いるデータのランダム性の担保しつつ高速化をおこなっている。この時の 1 ミニバッチに含まれるデータの個数をバッチサイズと呼び、ハイパーパラメータとして人為的に決定する。SGD の欠点は学習率の値と最適値周辺の形状によって収束に至るまで重みの値が振動する可能性があることである。学習率を小さくすることでこれを防ぐことは可能であるが、デメリットとして損失関数の収束が遅くなることが挙げられる。そのためこの問題を解決するために重みの変化によって学習率を変化させるアルゴリズムの必要性が叫ばれてきた。この解決策として現在では 2015 年に提唱された Adam(ADaptive Moment estimation)[87]が最も優れた性能を示しており広く用いられている。従前の SGD は勾配の更新単位が考慮されておらず、これまでの手法だと学習率だけに対して調整を行うような手法であった。そのため Adam では勾配の二乗平均と平均を 1 次モーメント v_t と 2 次モーメント s_t として考慮することで、パラメータ毎に適切なスケールで重みが更新されることを可能にした。Adam のアルゴリズムは以下の通りである。

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1)G \quad (3.13)$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2)G^2 \quad (3.14)$$

$$W_t = W_{t-1} - \eta \frac{v_t}{\sqrt{s_t + \epsilon}} \quad (3.15)$$

ここで G は重みの各重み更新における移動平均、 β_1 , β_2 はハイパーパラメータ、 ϵ は 0 除算を防ぐための微小値である。

3.4.2 誤差逆伝播法

誤差逆伝播法(backpropagation)[73]は重み最適化にて計算する勾配 $\nabla L(W)$ の項を効率的に計算する手法である。以下、説明のため 4 層ニューラルネットワークを仮定する。まず入力ベクトルを \mathbf{x} , 出力ベクトルを \mathbf{y} , 重み行列を W_i , バイアス $\mathbf{b}_i (i = 0, 1, 2)$, 活性化関数 $f(\mathbf{x})$ とする 4 層のニューラルネットワークによる順伝播は次式にて表される。

$$\mathbf{y} = f(W_2 \cdot f(W_1 \cdot f(W_0 \cdot \mathbf{x} + \mathbf{b}_0) + \mathbf{b}_1) + \mathbf{b}_2) \quad (3.16)$$

これを順伝播の経過を示す計算グラフで示すと図 3.3 になる。

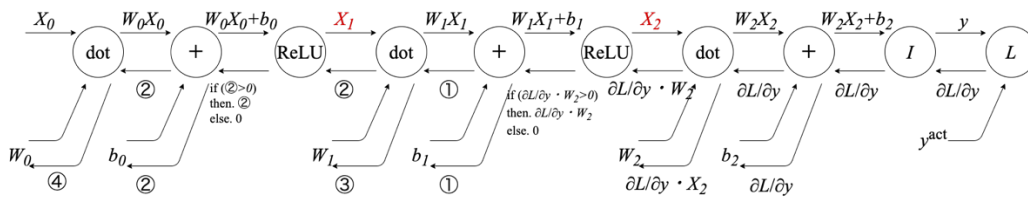


図 3.3 4 層ニューラルネットワークモデルにおける誤差逆伝播を示した計算グラフ。計算は右端の損失関数 L から左端の入力に向かって計算される。

ここで① $:= \partial L / \partial y \cdot W_2$, ② $:= \partial L / \partial y \cdot W_2 \cdot W_1$, ③ $:= \partial L / \partial y \cdot W_2 \cdot X_1$, ④ $:= \partial L / \partial y \cdot W_2 \cdot W_1 \cdot X_0$, $X_1 = W_0 X_0 + b_0$, $X_2 = W_1 \cdot X_1 + b_1$ である.

重みの更新式(3.10)中のそれぞれの勾配 $\nabla_w L$ はそれぞれの重み W_i ごとに異なり, それぞれ計算する必要がある. しかし $\nabla_w L$ の W は L の直接の引数ではないため, 出力の \mathbf{y} を介して合成関数として連鎖律

$$y = f(g(x)) \rightarrow \frac{dy}{dx} = \frac{dy}{dg} \frac{dg}{dx} \quad (3.17)$$

に基づいて計算する. ニューラルネットワークの訓練での演算は加算, 乗算, 非線形活性化, 損失関数で構成され全ての演算において初等関数による微分が定義されている.

- 加算ノード

$$z = x + y \rightarrow \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} = 1 \quad (3.18)$$

- 乗算ノード

$$z = xy \rightarrow \frac{\partial z}{\partial x} = y, \frac{\partial z}{\partial y} = x \quad (3.19)$$

- 活性化関数 ReLU のノード

$$z = \max(0, x) \rightarrow \frac{\partial z}{\partial x} = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3.20)$$

- 損失関数 MSE のノード

$$\begin{aligned} L(\mathbf{y}^{\text{train}}, \mathbf{y}^{\text{pred}}) &= \sum_i^N \frac{1}{N} (\mathbf{y}^{\text{train}} - \mathbf{y}^{\text{pred}})^2 \rightarrow \\ \frac{\partial L}{\partial \mathbf{y}^{\text{pred}}} &= \sum_i^N \frac{2}{N} (\mathbf{y}^{\text{train}} - \mathbf{y}^{\text{pred}}) \end{aligned} \quad (3.21)$$

これらを用いて, 例えば図 3.3 の全ての ReLU による活性化が全て 1 だった場合の W_1 の勾配を計算する.

$$\frac{dL}{dW_1} = \frac{dL}{dy} \cdot \frac{dy}{d(W_2 X_2)} \cdot \frac{d(W_2 X_2)}{dX_2} \cdot \frac{d(X_2)}{d\text{ReLU}} \cdot \frac{d(\text{ReLU})}{dW_1 X_1} \cdot \frac{dW_1 X_1}{dW_1} \quad (3.22)$$

$$= \sum_i^N \frac{2}{N} (\mathbf{y}^{\text{train}} - \mathbf{y}^{\text{pred}}) \cdot 1 \cdot W_2 \cdot 1 \cdot 1 \cdot X_1 \quad (3.23)$$

$$= \sum_i^N \frac{2}{N} (\mathbf{y}^{\text{train}} - \mathbf{y}^{\text{pred}}) \cdot W_2 \cdot X_1 \quad (3.24)$$

上記と同様に全ての W, b に対し実施し勾配を計算する.

3.4.3 ハイパーパラメータチューニング

特に人為的に決める必要があるハイパーパラメータの調整は重要であり，このパラメータによっては予測モデルの性能は大きく左右される．本研究では最も良い性能をもつ予測モデルを探索する手法としてグリッドサーチによるハイパーパラメータを行なった．グリッドサーチとは図 3.4 のようにハイパーパラメータの値をあらかじめいくつか決めておき，全てのパラメータの設定で計算し，最もよかったものを採用する方法である．他にランダムに値を振ることで偶然に良いモデルを見つけるランダムサーチが挙げられ，ハイパーパラメータ間の相関が少ない場合かつ探索空間内で最適なセッティングが 1 箇所に収束する場合に計算を高速化することが可能となる．

また事後分布に基づいて逐次的に最適なハイパーパラメータを探索するベイズ最適化[88]による探索も広く用いられており，一般的に探索空間内が複雑になりやすいニューラルネットワークにおいて大域的最適解を見つけやすいことが知られている．

他にも近年では多目的遺伝的アルゴリズム(Elitist Non-dominated Sorting Genetic Algorithm, NSGA-II)[89]などが様々な手法が提案され予測モデルの問題ごとに適したアルゴリズムで探索することが重要になってきている．

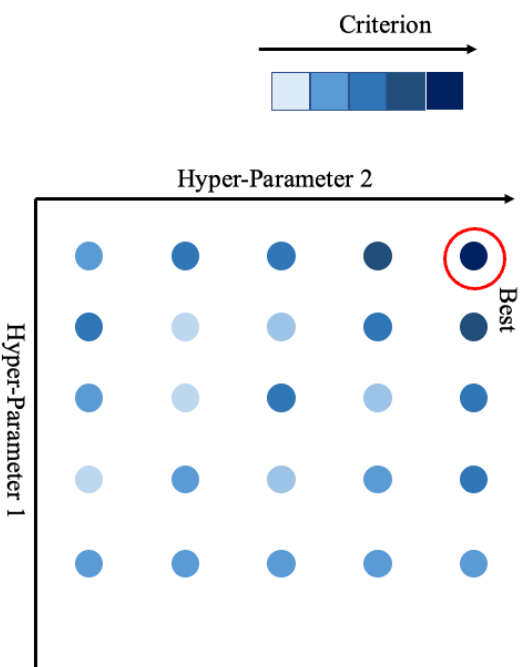


図 3.4 2 種類のハイパーパラメータとそれに対する評価の散布図. 色が濃いほど良い評価値であり，最も良い評価値の時のハイパーパラメータを採用する．

3.5 評価方法

機械学習における学習を最適化するためには学習率, エポック数などの人為的に決定されるハイパーパラメータやモデルの形状として隠れ層の数やそれぞれの隠れ層のノード数を最適化する必要がある. これらの最適化は以下に述べる統計的なパラメータによって評価し, 最もよい評価値を持つパターンモデルのニューラルネットワークの形状・ハイパーパラメータを選択し運用する必要がある. また評価方法は解決する問題によっても変わり, 本論文では回帰問題における評価方法について述べる.

3.5.1 決定係数

決定係数(coefficient of determination, R^2)は変数間の相関からモデルの当てはまりの良さを示す指標である. R^2 はさまざまな定義[90]が提案されているが, 本研究では `python` ライブラリの `scikit-learn` にて定義されたものを用いる. 計算式は図 3.1 の回帰モデルを用いると

$$R^2 = 1 - \frac{\sum_{i=0}^{N-1} (y_i - \hat{y}(x_i))^2}{\sum_{i=0}^{N-1} (y_i - \bar{y})^2} \quad (3.25)$$

であり, \bar{y} は \mathbf{y} の平均である. R^2 の値域は $-\infty < R^2 \leq 1$ であり, \mathbf{y} と $\hat{\mathbf{y}}(\mathbf{x})$ は $R^2=1$ では完全に一致, $R^2 \leq 0$ では無相関であり, 1 に近いほどモデルの当てはまりが良いことを示す.

3.5.2 学習曲線

学習曲線は縦軸を損失関数もしくは評価値, 横軸を学習回数としてプロットした曲線である. この曲線の目的は学習モデルの訓練が進むと損失が一定値に収束することを確認し, 収束した訓練データによる損失とテストデータによる損失のズレからモデルの汎化性能を評価することである. 回帰問題では図 3.5 のように縦軸には損失関数の値, もしくは誤差

(MSE の平方根である RMSE や絶対値平均誤差 MAE など), 横軸には学習回数であるエポック数を用い, 損失が一定値まで収束するか評価する. 理想は図 3.5(a), (b) のように訓練データによる誤差とテストデータによる誤差が一定値以内で収束もしくはテストデータが極小値をもつことであるが, 選択したハイパーパラメータによっては訓練データによる損失の収束が十分でないため汎化誤差が大きい過学習(under-fitting)や訓練データに過剰に適合しているためテスト誤差が大きくなる過学習(over-fitting)が想定される. また図 3.5(c) のような訓練データの収束に対しテストデータの収束が大きい場合をハイバリアンス, 訓練データ, テストデータの両方が大きい場合をハイバイアスと呼ぶ. これらを回避するためハイパーパラメータをより汎化誤差が小さくなるようチューニングする必要がある.

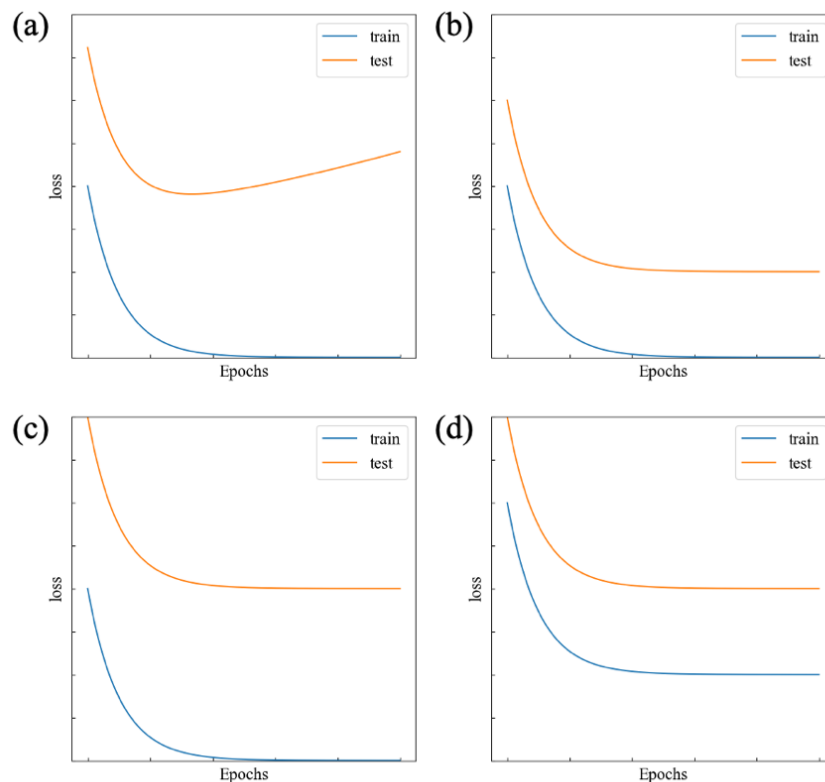


図 3.5 学習曲線の例. 一般的な学習ができている例として (a)テストデータに極小値を持つパターン(b)テストデータも収束を持つパターン. また(c)はハイバリアンス(d)ハイバイアスの状態の学習曲線である.

3.5.3 正解-予測データプロット

MSE, R^2 はモデルの当てはまりの良さを示す評価指標であるが, 回帰モデルによる誤差範囲から外れた「外れ値」に対する評価を行うことができない. そのためモデルの当てはまりと外れ値の存在を可視化するため, 図 3.6 のような横軸を正解データ, 縦軸を回帰モデルの予測データとした正解-予測データプロット(actual-predicted plot)を用いる. モデルの当てはまりに対する外れ値の評価は $y = x$ 線からの外れ具合と個数にて評価できるが, 許容範囲はデータの性質次第であるため定性的に評価する. しかし外れ値の具合と個数が多い場合はなどモデルを見直すなど評価の補助として用いることができる.

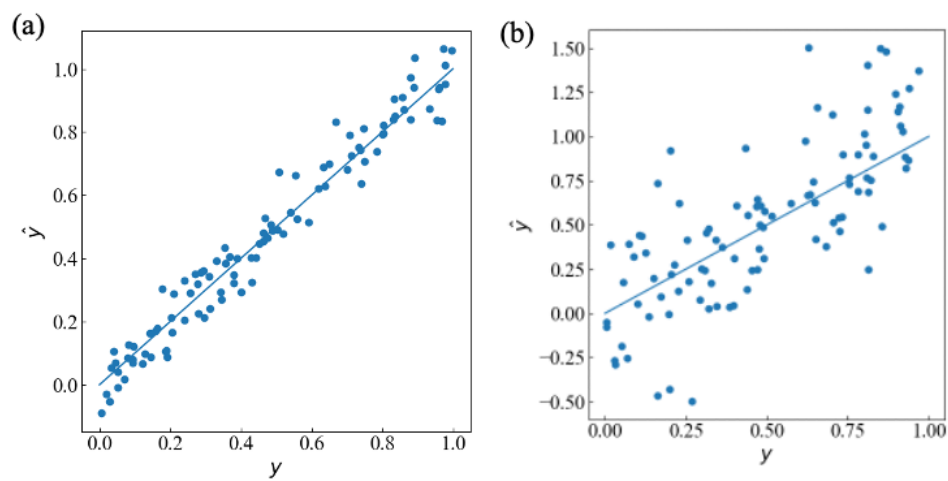


図3.6 正解-予測データプロットにおける(a)正しい予測ができている場合(b)できていない場合の例. データ分布の中心の線は理想的な予測を示す $y = \hat{y}$ の直線である.

3.6 計算フロー

以上の記述をフローチャートにしたものを図3.7に示す. まず学習・テストに用いるデータを取得したのちデータのスケールを揃えるなどの前処理を行う. 次に訓練データとテストデータに分割し, 訓練データにてモデルを最適化, テストデータにてモデルの汎化性能を計測する.

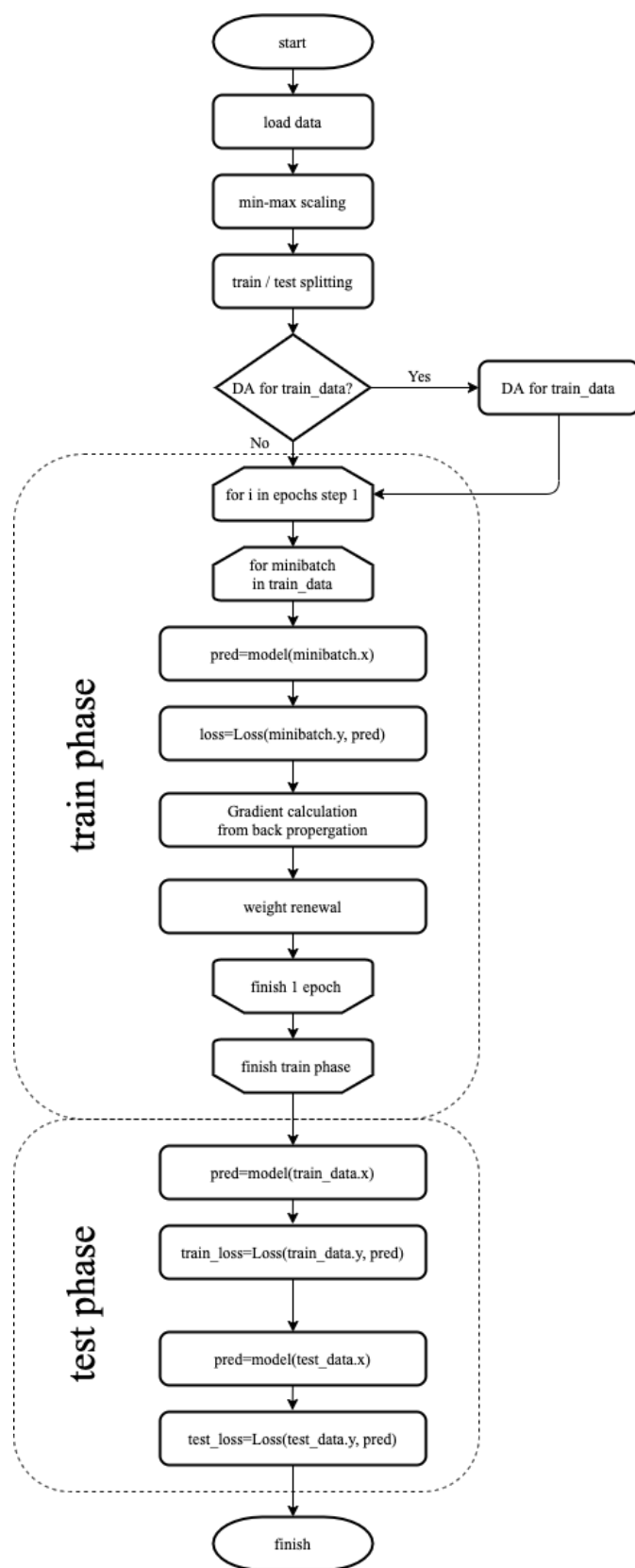


図 3.7 本研究における機械学習の訓練，テストの計算フロー．破線で囲まれた部分はそれぞれ訓練過程，テスト過程を示す．

3.7 ニューラルネットワークを用いた $U_{\text{eff}}, 1/\tau$ 予測手法

本研究では $U_{\text{eff}}, 1/\tau$ の組み合わせとそれに対応する第一原理誘電関数の関係を機械学習によって非線形回帰する手法を開発する。以下に図 3.8 に示す簡単な手続きについて説明する。まず $(U_{\text{eff}}, 1/\tau)$ の組み合わせを複数用意し、それぞれの第一原理誘電関数を計算する。次に機械学習手法であるニューラルネットワークに、第一原理誘電関数から $(U_{\text{eff}}, 1/\tau)$ を予測する様に学習を行い、誘電関数スペクトルからパラメータを推定する予測モデルを作成する（図 3.8①）。最後に得られた予測モデルに実験から観測された誘電関数を入力し（図 3.8②）、実験を再現すると期待される $(U_{\text{eff}}, 1/\tau)$ を推定する（図 3.8③）。本研究で機械学習手法にニューラルネットワークを選択したのは、同じく機械学習手法である決定木やサポートベクターマシンなどでは多値予測が困難なためである。ニューラルネットワークは基本的に行列による線形変換と非線形活性化によるものであるため、出力をベクトルとして扱うことが可能である。そのため本研究で予測する複数のパラメータ $(U_{\text{eff}}, 1/\tau)$ の予測に適している。

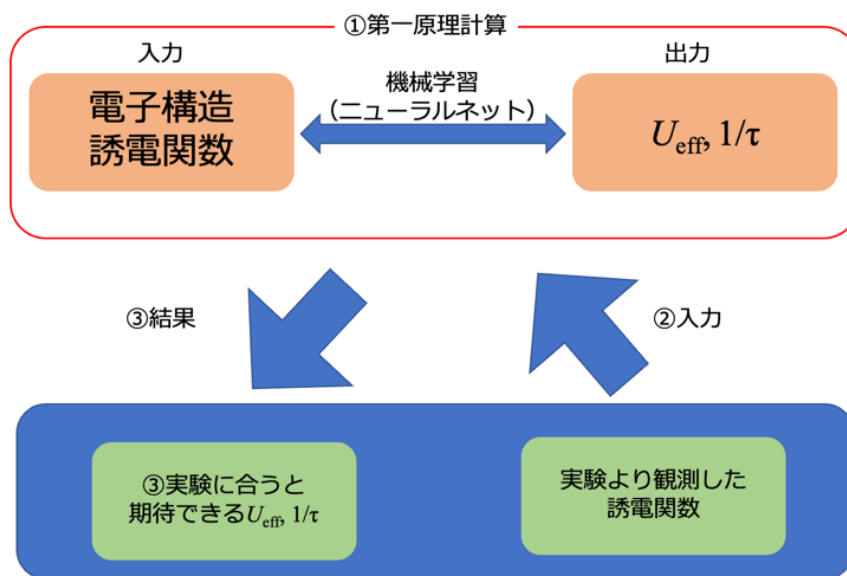


図 3.8 本研究で開発する誘電関数からパラメータを予測するシステムの概略図。計算の流れは図中の番号の順で行われる。

3.8 おわりに

本章では機械学習アルゴリズムの一つであるニューラルネットワークのアルゴリズムについて言及し、学習データに対する訓練方法と予測モデルの誤差逆伝播法に基づく重み最適化について概説した。また予測モデルの汎化性能を評価するため未知データを用いた統計的手法による定量的・定性的評価方法について示し、性能が良いモデルの持つべき性質を定義した。またこれら一連の計算フローについて示し、実際の計算に必要な手法を全て概説した。

第 4 章 NiO における機械学習手法の妥当性の検証

4.1 はじめに

酸化ニッケル NiO は典型的な $3d$ 遷移金属であり, 強相関効果を取り入れた第一原理計算手法の開発研究においてしばしばテストケースとして採用されている. 本章では先述のニューラルネットワークによる予測手法の妥当性の検証のため本予測手法を NiO の DFT+U 理論に基づく第一原理計算に適用する.

ニューラルネットワークによる学習のための訓練データ収集として NiO の第一原理計算による電子構造と誘電関数計算を実施する(4.2 節). 次にデータの前処理としてニューラルネットワークの学習に用いるデータを 4 種類の処理方法から選別し, 最も実験の誘電関数と近くなると期待できる前処理手法を検討する(4.3 節). 次に実際にニューラルネットワークによるモデルを構築し, ハイパーパラメータチューニングを実施(4.4 節), 最終的にもっとも性能がよかったモデル構成, ハイパーパラメータのセッティングで学習したモデルに対し, 実験スペクトルを入力することで最適な $U_{\text{eff}}, 1/\tau$ を予測する. 最後にそれらの値を用いた第一原理計算による電子構造, 誘電関数について実験と比較し, その妥当性を検証する.

4.2 第一原理計算による誘電関数データの収集

初めに, 予測モデルを訓練・テストを行うための学習データを収集する. NiO は図 4.1 に示すような格子定数 $a = 7.93 \text{ bohr}$ [91] の岩塩構造のプリミティブセル(格子ベクトル: $a_1 = a(1/2, 1/2, 0)$, $a_2 = a(0, 1/2, 1/2)$, $a_3 = a(1/2, 0, 1/2)$) 中に 4 つの原子(Ni : 2 原子, O : 2 原子) で構成され, (111) 方向に Ni 原子が反平行なスピン磁気モーメントを持つように配列した反強磁性材料である. 第一原理計算では Ni のマフィンティン球の半径を 2.2 bohr , O を 1.4 bohr とし, 基底関数(平面波)のカットオフを 3.9 bohr とした. 加えて, 図 4.2 の k 点数変化に対する全エネルギーの収束状態から k 点は $40 \times 40 \times 40$ を採用した. 次に学習データ収集のため上記条件を固定し U_{eff} をそれぞれ $0.0 \sim 9.9 \text{ eV}$ (100 パターン), 各 U_{eff} に対し $1/\tau$ を $0.1, 0.3, 0.5, 0.7, 0.9 \text{ eV}$ (5 パターン) の計 500 パターンについて計算を行った.

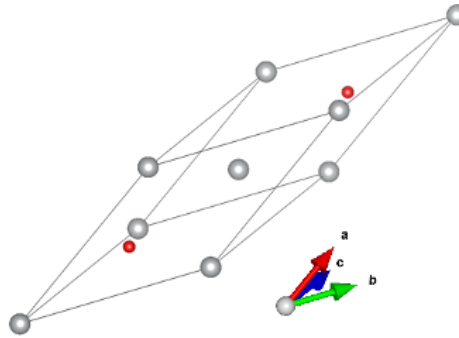


図 4.1 NiO のプリミティブセルのモデル. 灰色の球が Ni, 赤色の球が O を示す. また (111) 面に並行に Ni の磁気モーメントが反転する反強磁性を示す.

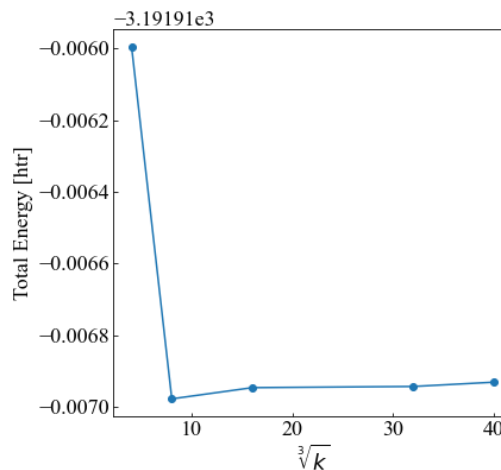


図 4.2 k 点に対する全エネルギーの変化. 横軸は k_x, k_y, k_z で同じ値をとっていることから 1 辺あたりの k 点の個数を示す.

4.3 機械学習に用いるデータ選択

4.2 で得られた第一原理誘電関数(光子エネルギーに対し 0.1 eV 刻みで離散化済み)についてデータの前処理として, まず最小を 0, 最大を 1 となるように正規化を施した. さらに機械学習の性能を向上させるため次の 4 つの前処理を実施し比較した.

- (a) 第一原理, 実験両方の誘電関数実部のみ取り出す.
- (b) 第一原理, 実験両方の誘電関数の実部, 虚部両方を取り出す (パターン A) .
- (c) 第一原理, 実験両方の実部, 虚部を取り出し, 第一原理の虚部に対してそれぞれの $U_{\text{eff}}, 1/\tau$ のパターンの電子構造計算から得られるバンドギャップ E_g より小さい光子エネルギー ω を全て 0 に統一する (すなわち $\epsilon_{xx}(\omega < E_g) = 0.0$, パターン B) .
- (d) 第一原理, 実験両方の実部と虚部を取り出し, 第一原理の虚部に対して, それぞれの $U_{\text{eff}}, 1/\tau$ のパターンの電子構造計算から得られる E_g より小さい光子エネルギー ω を, 全て光子エネルギーがバンドギャップの時の値に統一する (すなわち $\epsilon_{xx}(\omega < E_g) = \epsilon_{xx}(E_g)$, パターン C) .

図 4.3 は上記(a)~(d)による前処理を施した第一原理誘電関数と実験より得られた誘電関数[92]の MSE を計算し, その値を $U_{\text{eff}}, 1/\tau$ 依存性として可視化したものである. またそれぞれの図 4.3 の中にある星マークは全ての $U_{\text{eff}}, 1/\tau$ の組み合わせの中で最も MSE が小さかった組み合わせの位置であり, 詳細を表 4.1 に示す. (a)~(d)のパターンのうち最も MSE が小さかったのは(a)の実部のみ取り出したパターンであり, 値も他のパターンに対し 10^{-3} 小さくなった. これは Oppnear の理論[69]に基づく誘電関数計算にて発生する虚部の低光子エネルギー帯における計算エラーが MSE を押し上げているためと考えられる. また(c), (d)の処理では高光子エネルギー帯での誘電関数が負の値を示しており, 両者のバンドギャップでの値は全光子エネルギー帯の中で最低ではなかったため, 計算エラーをとりきれなかったと考えられる. したがって本手法において(a)誘電関数の実部のみの方法を用いて予測を行った. 本研究では誘電関数の実部, 虚部両方を最適化することが目的だが, 誘電関数の実部, 虚部の間にはクラマース・クローニツヒの関係が成り立つため, 第一原理誘電関数の実部が実験に対して最適化されていれば虚部も最適化されると仮定し, 実部のみを用いた最適化を行う.

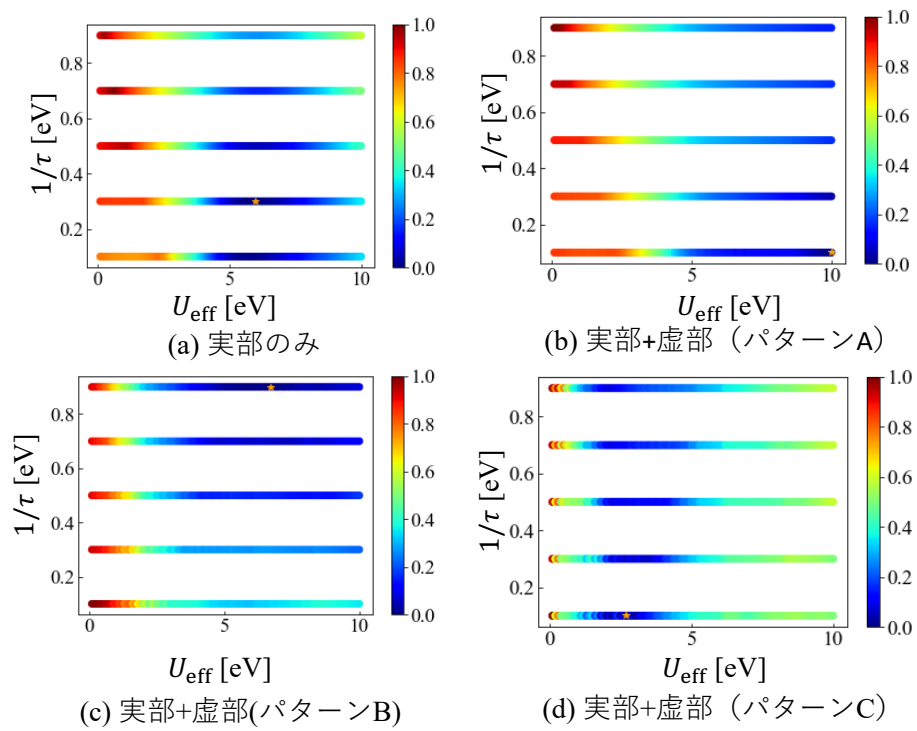


図 4.3 実験に対する第一原理誘電関数の MSE. (a)実部のみ計算に用いた場合, (b) 実部と虚部を用いた場合 (パターン A), (c)パターン B, (d)パターン C. MSE は全てのパターンを 0-1 で規格化している.

表 4.1 各前処理パターンにおける最小の MSE とそれに対応する U_{eff} , $1/\tau$ の比較. (a)-(d) は図 4.3 のものと一致する.

Method		MSE	U_{eff} [eV]	$1/\tau$ [eV]
(a) Realのみ		0.007067	6.0	0.3
Real + Imaginary	(b) パターンA	8.425666	10.0	0.1
	(c) パターンB	8.481092	6.7	0.9
	(d) パターンC	7.890614	2.7	0.1

4.4 学習・推論モデル

前処理として誘電関数のスペクトル高さ, U_{eff} , $1/\tau$ それぞれの範囲を 0 から 1 へ規格化された誘電関数実部を用いて予測モデルのモデル構成, ハイパーパラメータの最適化を行った. この予測モデルは入力層を 200 次元, 出力層を 2 次元(U_{eff} , $1/\tau$)で固定し, チューニング対象となるパラメータを, 学習率, 隠れ層の数とそれぞれのノード数, エポック数とし, これらのチューニングを行った.

まず図 4.4 にテストモデルとして隠れ層を 1 層のみ 100 ノード, 100 エポックの計算を, 学習率を 10^{-2} , 10^{-3} , 10^{-4} で変化させた時のテストにおける学習曲線を示す. 計算を行うにあたり再現性を担保するため, 同じハイパーパラメータの設定に対し乱数シードを 10 パターン用意しそれぞれ計算を行い, 最終的に平均を取ることでその設定に対する結果とした. 全てのパターンにおいて学習曲線はエポック数が大きくなるほど減少しており, 正しく学習が進んでいることがわかる. また学習率間を比較して 10^{-3} の学習曲線が最も MSE が小さくなった. これは, 10^{-2} の場合は重み更新における学習率が高すぎることによる重み更新の振動の発生のため, 10^{-4} の場合は逆に学習率が低すぎることによる未学習が発生しているためと考えられる. よって以上のことから 10^{-3} が最適であると判断し以後の学習ではこの値を用いる.

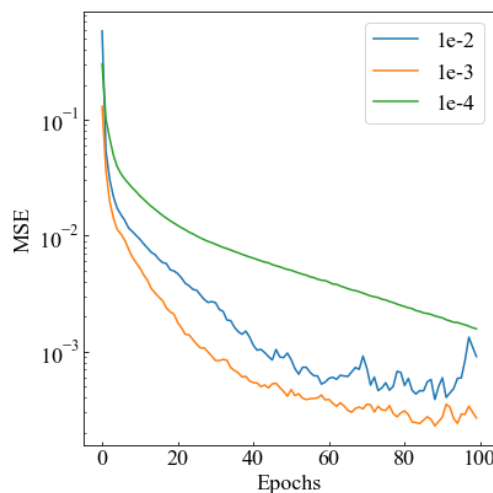


図 4.4 学習率ごとのテスト誤差の収束の比較. 各曲線は学習率が 10^{-2} (青色), 10^{-3} (橙色), 10^{-4} (緑色)を示す. 縦軸は変化を強調するため対数とした.

次にニューラルネットワークの隠れ層の数と各ノードの個数についてチューニングを行った. 隠れ層が 1 層のみと 2 層の場合に対して実施し, それぞれ 50, 100, 150 ノードの組み合わせにて実施した. それらの時のテストデータによる MSE について比較した. その結果図 4.5 のように 2 層構成で 1 層目が 150 ノード, 2 層目が 100 ノードの時に最も MSE が小

さいことがわかった。また層やノードによる比較を行ったところ傾向を確認することができず、ノード数の変化に対して複雑に変化することがわかった。

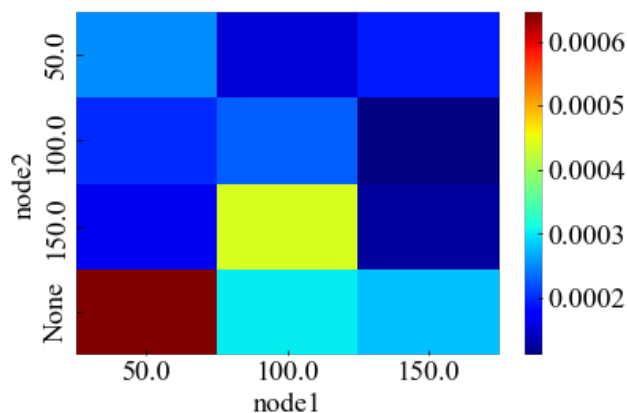


図 4.5 ノードの隠れ層の組み合わせによるテストデータの MSE の比較。横軸を 1 層目のノード数, 縦軸を 2 層目のノード数を示す。縦軸の None は隠れ層が 1 層の時のものを示す。

最後に隠れ層 1 層目を 150 ノード, 2 層目を 100 ノードに設定し, 130 エポックまで計算を実施し, エポックに対するモデルの性能を比較した。その結果 50 エポックまでは訓練, テスト両方とも MSE は単調に減少したが, それ以降は振動しつつも一定に収束した。この中で最もテスト MSE に対して値が小さかった 100 エポックを採用した。また図 4.8 にこの時の正解・予測データプロットを示す。(a)の U_{eff} に対する予測では訓練データ, テストデータ両方に対する予測はほぼ $y = x$ 線上に分布し, 極めて良い予測をしていることがわかる。また決定係数 R^2 も 0.999 以上と極めて高く, 良い予測が行われていることがわかる。一方(b)の $1/\tau$ は訓練データに対してテストデータがわずかに予測にばらつきがあるもののほとんどの予測が $y = x$ 線上に分布しており良い予測をおこなっていることがわかる。また R^2 も U_{eff} に劣るものの 0.998 以上と良い予測をおこなっており, 全体として良い予測を行うモデルを構築することに成功した。 $1/\tau$ の R^2 が U_{eff} に対して劣る理由は U_{eff} が誘電関数のピーク位置に対する予測であることにに対し, $1/\tau$ はピーク周辺の分散に対する予測であり, ニューラルネットワークがことなる 2 つの予測に対し U_{eff} を優先してフィッティングしたためと考えられる。しかし両者とも R^2 は 0.99 以上を示していることから本予測モデルの精度は十分であると考えられる。

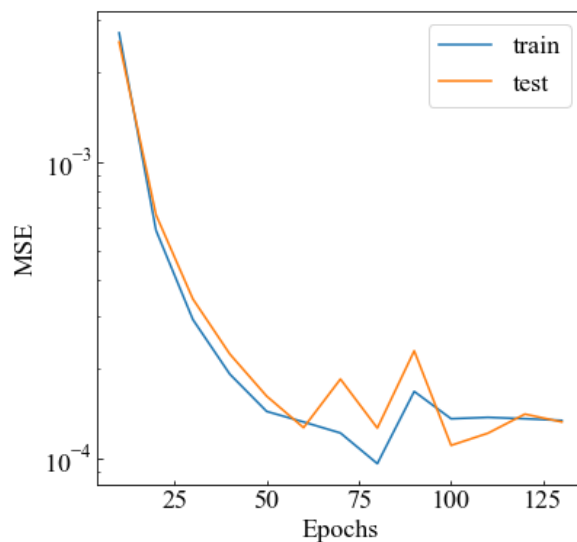


図 4.6 エポックごとの損失関数の比較. 青線が訓練データによる MSE, 橙線がテストデータによる MSE を示す. 変化を強調するため縦軸は対数目盛で表示した.

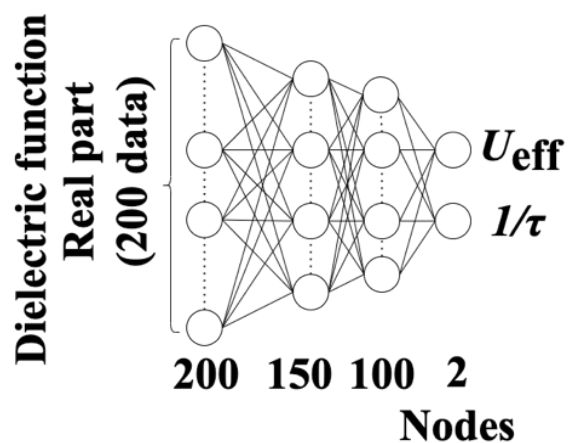


図 4.7 本章の計算で用いた NN 予測モデルの模式図. 各○がノードを示しており, 数が膨大なため破線で省略して描写した. また線は各ノード間の接続を表しており, それぞれの線で重みを持つ. 計算は左端の誘電関数入力部から始まり右端に至るまで Affine 変換と ReLU による非線形活性化を行う.

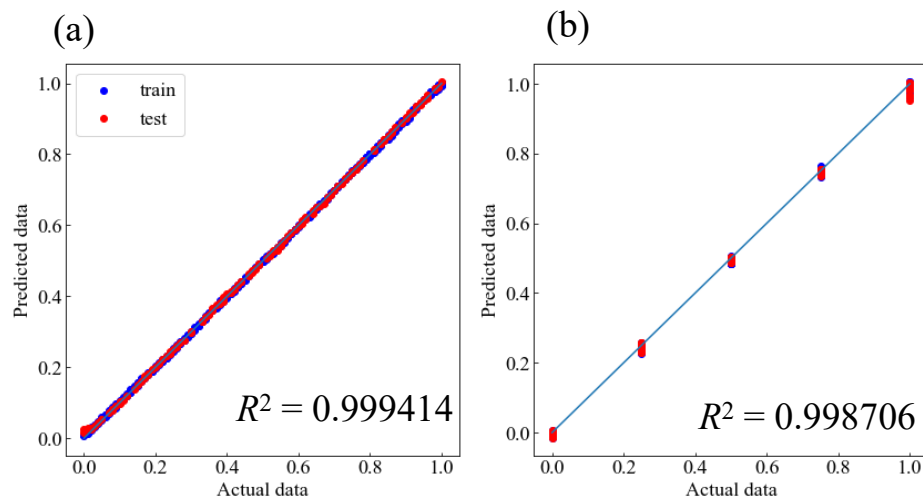


図 4.8 (a) U_{eff} , (b) $1/\tau$ に関する正解データ-モデル予測データプロット. 青色のプロットは訓練データを用いて行なった予測, 赤色のプロットはテストデータを用いて行なった予測を示す. 中央の青線は $y = x$ を示す. U_{eff} , $1/\tau$ それぞれのテストデータにおける R^2 を右下に示す. また U_{eff} と $1/\tau$ の値の範囲が異なるためそれぞれ 0-1 の範囲で規格化した.

4.5 第一原理計算との比較

4.4 にて構築した予測モデルを用い、実験から得られた誘電関数スペクトルを入力すると、 $U_{\text{eff}} = 5.52\text{eV}$, $1/\tau = 0.551\text{eV}$ を得た。この結果を元に第一原理計算による電子構造、誘電関数を計算した。

まず U_{eff} について先行研究と比較した。先行研究[36, 93-97]にて算出された U_{eff} は 5eV から 7eV の範囲で算出されており、本研究から得られた値もその範囲で得られた。

表 4.2 NiO の Ni 原子に対する有効オンサイトクーロン相互作用 $U_{\text{eff}} (= U - J)$ と $1/\tau$ の先行研究との比較。第 1, 2 列は、それぞれ計算時に使われた基底関数と計算手法を示している。

基底	手法	U_{eff}	$1/\tau$	Ref.
LAPW ^a	機械学習	5.52	0.551	This work
PW ^b	Linear Response	5.77	-	[36]
LMTO ^c	Constrained LDA	7.05	-	[93]
LCAO ^d	Constrained LDA	5.1	-	[94]
PAW ^e	cRPA	6.4	-	[95]
		6.0	-	[96]
MLWF ^f	cRPA	5.6	-	[97]

a: linearized augmented plane wave

b: linear muffin-tin orbitals

c: linear combination of atomic orbitals

d: projector-augmented wave

e: Maximally localized Wannier functions

f: ...

次にこれらの値を用いて第一原理誘電関数を計算した。その結果を図 4.9 に示す。第一原理計算より得られる誘電関数に対する U_{eff} と $1/\tau$ はそれぞれ主にスペクトルのピーク位置とピーク周辺における分散に大きく寄与する。まず、機械学習の訓練に用いた誘電関数実部に関して実験の誘電関数と比較すると、最大ピーク位置は 3.7eV であり実験(3.9eV)と比較して良い一致を示した。またピーク周辺の分散は実験と比べて過大評価されており、定量的な予測には至らなかった。

これは、予測モデルにおいて $1/\tau$ の予測の重要度が U_{eff} の予測に比べて低く判断したためと考えられる。 U_{eff} の場合、値は最大ピーク位置に大きく依存し、その変化はピーク周辺の分散に寄与する $1/\tau$ よりも変化の幅が大きい。そのため予測モデルは U_{eff} に対する変化を重要視したと考えられる。

次に誘電関数虚部について実験の誘電関数と比較すると最大ピーク位置は 4.5eV であり実験(4.2eV)と比較して良い一致を示した。またピーク位置の分散に関して良い一致を示し

ており、良い予測ができていますと結論づけた。

以上のことから本手法による誘電関数スペクトルの実験の再現は最大ピーク位置については定量的に、ピーク周辺の分散は定性的に一致することができ、本手法の妥当性を示すことに成功した。

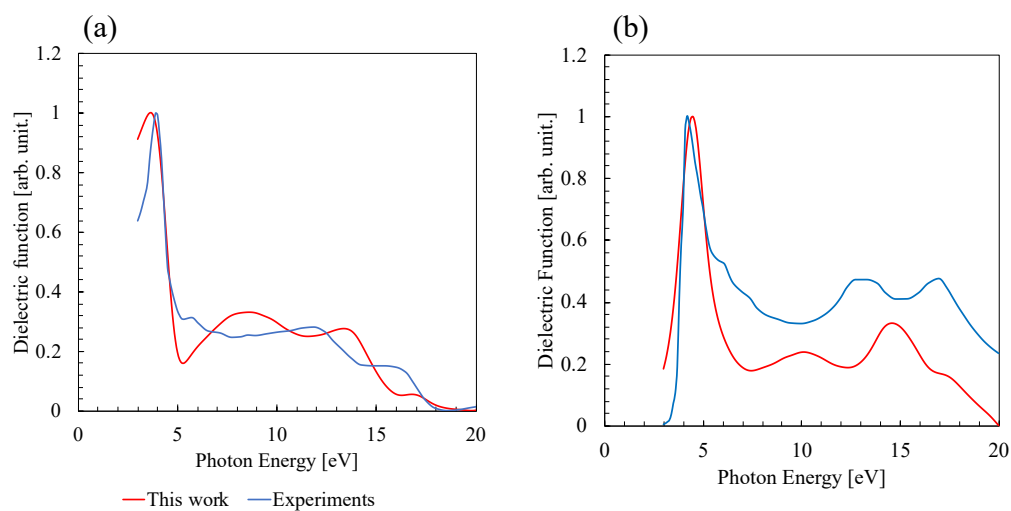


図 4.9 第一原理から計算した誘電関数 ϵ_{xx} (赤線) と実験から得られた誘電関数[92] (青線) の(a)実部, (b)虚部. それぞれのスペクトルは 0 から 1 の範囲で規格化してある.

4.6 まとめ

本章では NiO に関する予測モデルの構築のためのデータ選択とハイパーパラメータチューニングについて述べて、構築した予測モデルで $(U_{\text{eff}}, 1/\tau)$ の最適値を予測した。またそれらの値を用いて第一原理計算を行い、得られた誘電関数と実験から観測された誘電関数を比較し、本手法の妥当性を検討した。

まず 4.2 節ではランダムに選択した $(U_{\text{eff}}, 1/\tau)$ の組における第一原理計算を行い、誘電関数を計算するに至るまでの計算条件について詳しく述べた。

続いて 4.3 節では得られた $(U_{\text{eff}}, 1/\tau)$ ・誘電関数対角項 ϵ_{xx} データセットを用い、第一原理計算から得られた誘電関数を実験と比較することでどのデータを選択すると最も MSE が小さくなるか検討した。その結果誘電関数の実部のみを用いた場合が極めて MSE が小さくなり、予測モデルの訓練にはこのデータを用いることに決定した。またこれらの手法ごとに MSE が大きく異なる点について議論した。

4.4 節ではニューラルネットワークを構築し学習率、隠れ層のノード数と層数、エポック数についてハイパーパラメータチューニングを行なった。その結果、学習率 10^{-3} 、隠れ層が 2 層で 1 層目が 150 ノード、2 層目が 100 ノード、エポック数が 100 の場合が最適であることが判明した。またこれによるモデルの性能評価（正解・予測プロット、 R^2 ）も行い、予測モデルの性能についても議論した。

最後に 4.5 節では実験から得られた誘電関数を入力し、 $(U_{\text{eff}}, 1/\tau)$ の最適値を予測した。またこれらの値を用いた第一原理計算から誘電関数を計算し、実験と比較を行なった。第一原理計算から計算した誘電関数の形状は実験とは定性的に一致し、最大ピーク位置は実験と極めて近い値になったが、 $1/\tau$ は過大評価された。以上のことについて議論を行なった。

第 5 章 YIG におけるパラメータ予測とデータ拡張による性能向上

5.1 はじめに

前章では NiO を用い、本研究で採用した機械学習によるパラメータ予測手法の妥当性を示した。本章ではこの手法を YIG に対して適用し、前章と同じく第一原理計算によるデータ収集(5.2 節)について述べ、これらのデータを用いておこなった予測モデルの性能について考察する(5.3 節)。続いてデータ数不足を補うためガウスノイズによるデータ拡張を実施する。これは第一原理計算より得られた誘電関数に対し乱数によるノイズを付加することでオリジナルデータに類似した新しいデータを生成し、それらを訓練データに含めることでデータ総数を増加させる手法である。本研究では最適なデータ拡張の設定を決定するため、3 種類のハイパーパラメータを設定しそれについてグリッドサーチを行い、変化に対する性能を比較する(5.4 節)。またデータ拡張がどのように予測モデルに寄与するか明らかにするため、正則化手法である **weight decay** による予測モデルの予測結果と比較する(5.5 節)。最後に予測モデルに実験から観測された誘電関数を入力して得られた U_{eff} , $1/\tau$ を用いて第一原理計算を行い、得られた電子構造、誘電関数を実験と比較した。また同様に磁気光学カー効果についても第一原理計算から得られた結果と実験結果を比較し考察する(5.6 節)。

5.2 第一原理データ収集

初めに、予測モデルを訓練・テストを行うための学習データを収集する。YIG は前述の通りイットリウム(Y), 鉄(Fe), 酸素からなるガーネットが構造を持つ酸化物である。図のような立方晶としてのユニットセルは 160 個という膨大な原子数を保つため、計算コストの面からユニットセル内の原子数が 80 個となる bcc のプリミティブセルを用いて計算をおこなった。実験より得られた格子定数 a は 23.387 bohr[98] であり、格子ベクトルはそれぞれ $a_1 = a(-1/2, 1/2, 1/2)$, $a_2 = a(1/2, -1/2, 1/2)$, $a_3 = a(1/2, 1/2, -1/2)$ である。ユニットセル内において、それぞれ Y イオンが十二面体配位, Fe イオンが八面体配位(Fe^0), 四面体配位(Fe^{T})を占有し、O イオンがその周りを配位しており、Fe イオンの d 電子軌道に対して U_{eff} による計算補正 ($U_{\text{eff}}^0, U_{\text{eff}}^{\text{T}}$)をおこなった。また Y 原子, Fe 原子, O 原子の muffin-tin 球の半径をそれぞれ 2.5 bohr, 2.0 bohr, 1.5 bohr とした。LAPW 基底関数のカットオフ値として、 $|\mathbf{k} + \mathbf{G}| \leq 3.6 \text{ bohr}^{-1}$ を用いており、muffin-tin 球内での波動関数、電子密度及びスピン密度、ポテンシャルの角運動量展開はそれぞれ $\ell = 8(\text{Y, Fe})$ と $\ell = 6(\text{O})$ を用いた。交換相関関数には一般化勾配近似 (GGA) を用いた。また第一 Brillouin Zone 内で、 $2 \times 2 \times 2$ の \mathbf{k} 点を用いて計算を行った。また得られた電子構造から線形応答理論に基づいて YIG の誘電関数及び磁気光学カー効果について計算をおこなった。ここで光学伝導度の計算における緩和時定数 $1/\tau$ は可変とし、それぞれのスペクトルは光子エネルギーに対し 0.1 eV 刻みで計算を行い、それぞれ光子エネルギーの変化に対し 50 個(0.1 ~ 5 eV)のデータを取得した。

これら一連の計算を ($U_{\text{eff}}^0, U_{\text{eff}}^{\text{T}}, 1/\tau$) を様々な組み合わせについて実施した。 ($U_{\text{eff}}^0, U_{\text{eff}}^{\text{T}}$) の組み合わせは 0~10 eV の範囲で 35 パターン計算した。また $1/\tau$ はそれぞれの ($U_{\text{eff}}^0, U_{\text{eff}}^{\text{T}}$) に対し 0.01, 0.03, 0.05, 0.07, 0.09 の 5 パターンを計算し合計 175 パターンの ($U_{\text{eff}}^0, U_{\text{eff}}^{\text{T}}, 1/\tau$)-誘電関数データセットを構築した。

5.3 学習・推論モデル

得られた $(U_{\text{eff}}^0, U_{\text{eff}}^T, 1/\tau)$ -誘電関数データセットを、前処理として誘電関数のスペクトル高さ、 U_{eff} , $1/\tau$ それぞれの範囲を 0 から 1 へ規格化し、これらをもとに図 5.1 のようなニューラルネットワークによる予測モデルを構築した。入力層は NiO の場合と同じく誘電関数対角項 ϵ_{xx} 実部の 50 ノード、出力層は $(U_{\text{eff}}^0, U_{\text{eff}}^T, 1/\tau)$ の 3 ノードである。また隠れ層は各 40 ノードの 3 層構造であり、入力層、隠れ層からの重みによる線形変換後は ReLU 関数、最終隠れ層のあとには恒等変換による活性化関数を用いる。またハイパーパラメータとして学習率は 10^{-3} 、エポック数は 2000 を採用した。損失関数には MSE を用いた。

図 5.2 にこの時の学習曲線の収束を示す。学習データを訓練データ(100 個)とテストデータ(75 個)に分割した。また乱数による偶然性の影響を軽減するため、ニューラルネットワークの計算全般に渡って用いる乱数シードを 10 パターン計算し、それらの結果を平均した。訓練時における収束は 10^{-3} まで収束し、正しく学習が行われていることわかる。またテストデータによる収束は 2000 エポックまで単調減少しておりその後 10^{-2} で収束した。

2000 エポックまで学習した時の正解・予測データプロットと決定係数 R^2 を図 5.3 に示す。図 5.3(a), (b)の U_{eff} について青色のプロットである訓練データはほぼ $y = x$ に分布しており良い予測ができていることがわかり、赤色のプロットであるテストデータは外れ値もいくつか存在するが、大部分のデータにて正しい予想が行われている。またこれらの R^2 は 0.95 以上を示しており、 U_{eff}^0 , U_{eff}^T に関して定性的・定量的両面で良い予測が行われていることが窺える。しかし図 5.3(c)の $1/\tau$ は U_{eff} と比較して、訓練、テストデータの両方で予測のばらつきがみられる。このばらつきは予測範囲の端の方に行くほど大きくなり、 $1/\tau$ が 0.01, 0.09 の場合と 0.03, 0.05, 0.07 の場合を比較すると前者の方がテストデータの予測にて大きなばらつきがあることがわかる。また $1/\tau$ の R^2 は 0.705 と U_{eff} に比べて小さくなっている。そのためこの予測モデルは U_{eff} に対してよい予測をするものの、 $1/\tau$ に対しては性能が高くないことがわかった。この原因は U_{eff} が誘電関数のピーク位置に対する予測であることに対し、 $1/\tau$ は誘電関数スペクトルのピーク周辺の分散に対する予測であることから、予測する特徴が異なること、またピーク位置に対してピーク周辺の分散は変化が小さいため、予測モデルが特徴を捉えきれなかったためと考えられる。また、本研究で用いた誘電関数のデータ数(175 個)は NiO(500 個)に比べて小さく、データ数不足も性能が向上しない原因として考えられる。以上の観点から $1/\tau$ の性能を向上させるためには、 $1/\tau$ に関してスペクトル形状を U_{eff} と同じ程度大きな依存性を持たせる、データ数を増加させることが重要となる。しかし前者は、 $1/\tau$ によって第一原理計算の誘電関数を加工して学習を行うと、実験より得られた誘電関数を入力する際正しく予測することができないと考えられる。そこで改善策として後者の訓練データ数を増加させることで対処することで改善を図る。次節ではデータ数を増加させるためにおこなったデータ拡張を実施した結果について述べる。

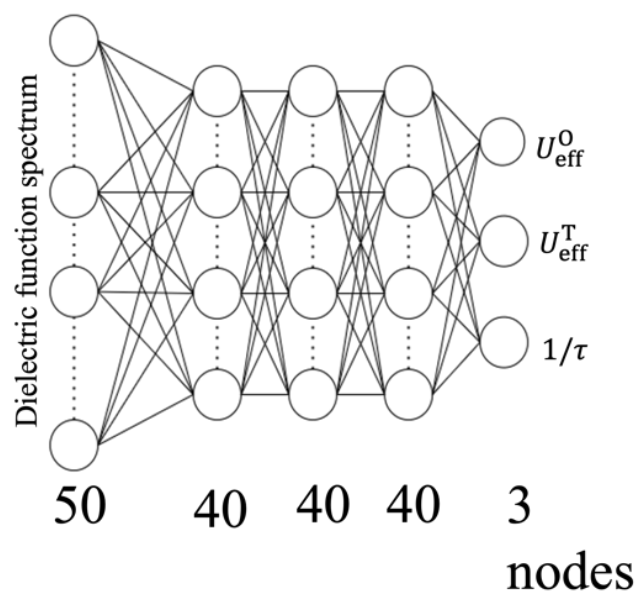


図 5.1 本章の計算で用いた NN 予測モデルの模式図. 各○がノードを示しており, 数が膨大なため破線で省略して描写した. また線は各ノード間の接続を表しており, それぞれの線で重みを持つ. 計算は左端の誘電関数入力部から始まり右端に至るまで Affine 変換と ReLU による非線形活性化を行う.

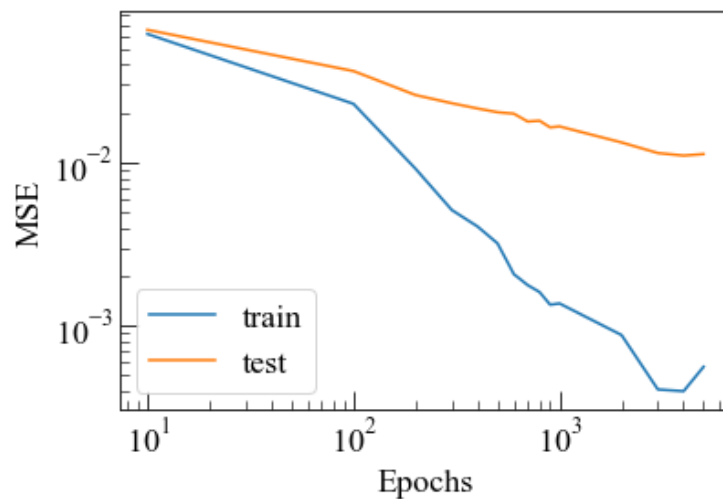


図 5.2 本章の計算における学習曲線. 青色の線が訓練における MSE, 橙色がテストにおける MSE を示す. 曲線の変化を強調するため両対数で描写している.

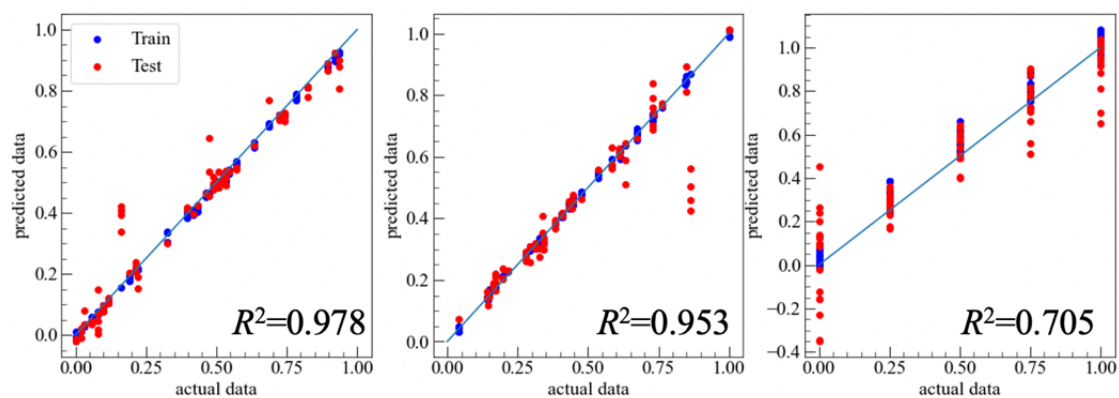


図 5.3 2000 エポックの時の正解-予測プロット. 青色の点が訓練データによる予測, 赤色の点がテストデータによる予測. 中央の線は $y = x$ の直線である. U_{eff} と $1/\tau$ でデータ範囲が異なるため, 0-1 の間で規格化している. また各グラフの右下にテストデータにおける R^2 を示した.

5.4 ガウシアンノイズによるデータ生成と拡張パラメータの選定

5.3 節では 175 データによる学習を行い、その結果予測モデルは U_{eff}^O と U_{eff}^T に関して良い予測をしたが、 $1/\tau$ に関して U_{eff}^O と U_{eff}^T よりも性能が劣った。この原因としてデータ数不足を挙げ、その対処法としてデータ拡張法を実施する。

本節では式(3.6)によるガウシアンノイズによるデータ拡張を実施する。本手法においてはガウシアンノイズの分散 σ_G^2 、訓練におけるエポック数、データ拡張で増やすデータの倍率 n をハイパーパラメータとして変化させ、最も良い性能を示す組み合わせを探索するとともに、変化に対する依存性を考察する。なお倍率 n から実際の訓練データへの変換は次式で表される。

$$N_{\text{train}} = N_{\text{original}} + nN_{\text{original}} \quad (5.1)$$

なお N_{train} は訓練データの総数、 N_{original} はオリジナルデータの個数である。

まずオリジナルの訓練データを 100 個に固定したまま、 $\sigma_G^2(10^{-8}, 10^{-7}, \dots, 10^{-1})$ 、エポック数(100, 200, 1000, 2000)、 $n(1, \dots, 200)$ についてそれぞれ変化させたところ、最も性能が良かった組み合わせはそれぞれ 10^{-3} , 2000, 3 であった。この時の正解-予測プロットと R^2 を図 5.4 に示す。まず R^2 をデータ拡張前である図 5.4(a)-(c)とデータ拡張後である図 5.4(d)-(f)を比べると、 U_{eff}^O , U_{eff}^T , $1/\tau$ のどの場合においても向上が見られた。また正解-予測プロットにおいては $1/\tau$ では大きな変化は見られなかったが、 U_{eff}^O , U_{eff}^T では外れ値の個数、 $y = x$ 線上からのずれも小さくなっていることがわかる。以上のことからデータ拡張は予測モデルに対し性能が向上するように寄与していることがわかった。また $1/\tau$ に関して R^2 は 10% 程度向上していることから、データ拡張なしの場合に比べて良い予測ができるようになっていたことがわかった。

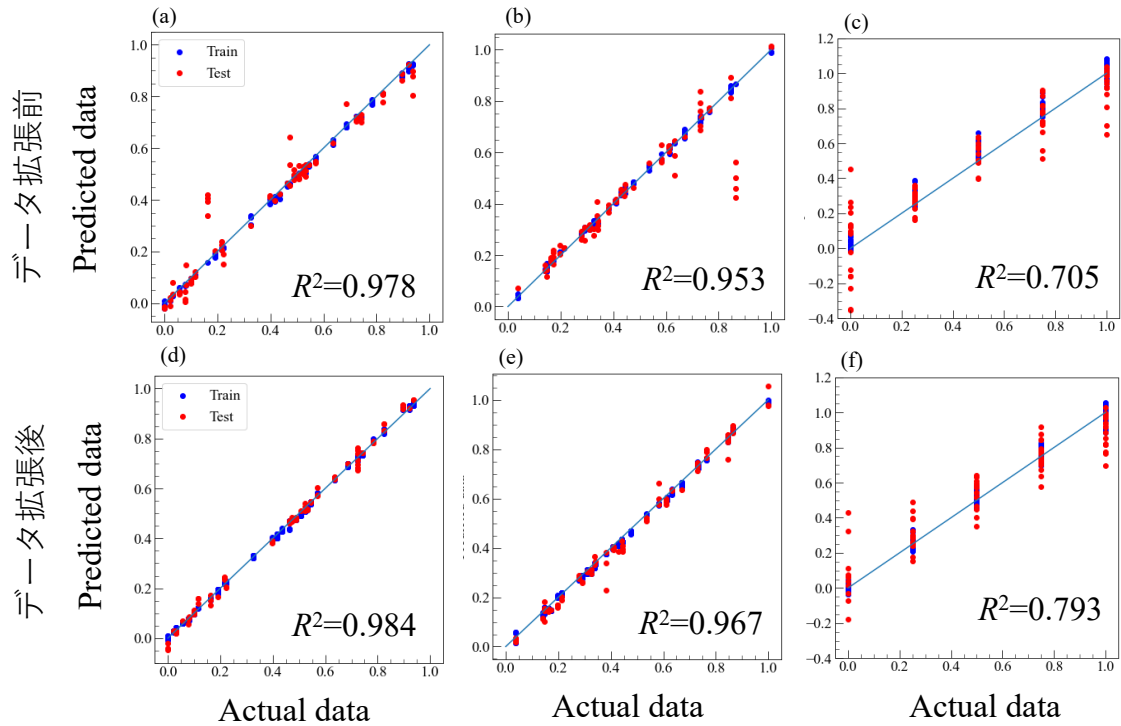


図 5.4 (a)-(c)データ拡張前の U_{eff}^O , U_{eff}^T , $1/\tau$ の正解-予測プロットとテストデータにおける R^2 (図再掲). (d-f)データ拡張後の U_{eff}^O , U_{eff}^T , $1/\tau$ の正解-予測プロットとテストデータにおける R^2 (図 5.3 再掲). 青色の点が訓練データによる予測, 赤色の点がテストデータによる予測. 中央の線は $y = x$ の直線である. U_{eff} と $1/\tau$ でデータ範囲が異なるため, 0-1 の間で規格化している.

また σ_G^2 , 訓練におけるエポック数, データ拡張で増やすデータの倍率 n について先述の最高性能を示す組み合わせを通してそれぞれのパラメータを変化させた.

まず σ_G^2 特性としてエポック数を 2000, 拡張倍率を 3 倍に固定したまま, σ_G^2 を 10^{-8} から 10^{-1} まで 1 桁ずつ桁刻みで変化させた. この時の U_{eff}^O , U_{eff}^T , $1/\tau$ のテストデータにおける R^2 の変化を図 5.5 に示す. この結果より σ_G^2 を 10^{-2} 以上にすると, それよりも小さい σ_G^2 に比べて R^2 は急激に悪化した. これはオリジナルデータに対してノイズの大きさが過剰に大きくなったことから, 正しく訓練・予測ができなくなったと考えられる. また U_{eff}^O , U_{eff}^T , $1/\tau$ の間で比較するとこの影響は $1/\tau$ において顕著であり, σ_G^2 が 10^{-2} ではデータ拡張前よりも性能が悪化しており, データ拡張以前で予測が困難であったもののほどノイズによる擾乱の影響は大きくなると考えられる.

次にエポック数特性として σ_G^2 を 10^{-3} , 拡張倍率を 3 倍のままエポック数を 10, 100, 1000, 2000 と変化させた. 本特性を測定する目的は, データ数が増加したことによりエポック数が少ない段階で一定に収束するか確認するためである. これはデータ拡張により 1 エポッ

クあたりのミニバッチの個数が増えたため重み更新の回数がデータ拡張なしの場合に比べて多くなる。そのため早く収束するか変化を確認する必要がある。この時の U_{eff}^O , U_{eff}^T , $1/\tau$ のテストデータにおける R^2 の変化を図 5.6 に示す。この変化は U_{eff}^O , U_{eff}^T においてはほぼ平坦であり、エポックにより大きく性能が変化しないことがわかった。しかし $1/\tau$ に関してエポックが増加するほど R^2 が増加し続けており、性能が改善していることがわかる。これは学習回数が増えることでアンダーフィッティングが解消され続けていると考えられ、このことはデータ拡張の有無に関わらず同じ状況である。そのため純粋に学習回数が増えたことによって予測モデルの性能が向上したと考えられ、この特性はデータ拡張による寄与が少ないと考えられる。

最後に n 特性として σ_G^2 を 10^{-3} , 2000 エポックのまま n を 1~200 の間で変化させた。この時の U_{eff}^O , U_{eff}^T , $1/\tau$ のテストデータにおける R^2 の変化を図 5.7 に示す。この時 $n=10$ 以下ではデータ拡張の効果によって R^2 が向上されており、ノイズによる擾乱が予測モデルのロバスト性を向上させていると考えられる。しかし n がそれ以上の場合ではデータ拡張以前と比べると R^2 が悪くなっており、これはデータ拡張によって増やされるノイズ入りデータ数が過剰になったためと考えられる。そのためオリジナルの情報が訓練に反映されなくなり結果としてテストデータによる R^2 も悪くなったと考えられる。

以上 3 特性からデータ拡張による予測モデルの性能改善にはガウシアンノイズの分散 σ_G^2 と拡張倍率 n が大きく寄与していることが明らかとなった。またこれらの値は大きいほど良いというわけではなく、それぞれの予測モデルの性能（本研究ではテストデータの R^2 ）を見ながら調整する必要がある。訓練データ、予測モデルに合わせて最適化する必要がある。

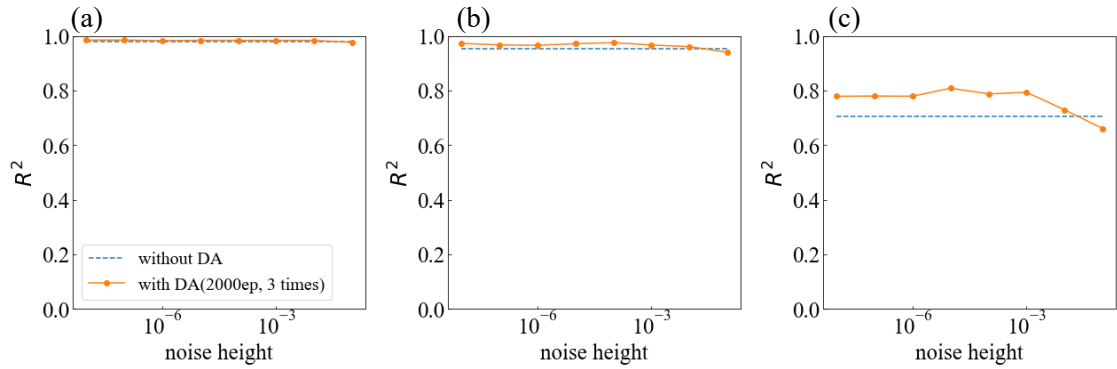


図 5.5 (a) U_{eff}^0 , (b) U_{eff}^T , (c) $1/\tau$ のテストデータにおける R^2 の σ_G^2 特性. 橙色のプロットがデータ拡張(2000 エポック, 3 倍拡張)した時の特性. 青色の破線はデータ拡張なしの時のテストデータにおける R^2 を示す.

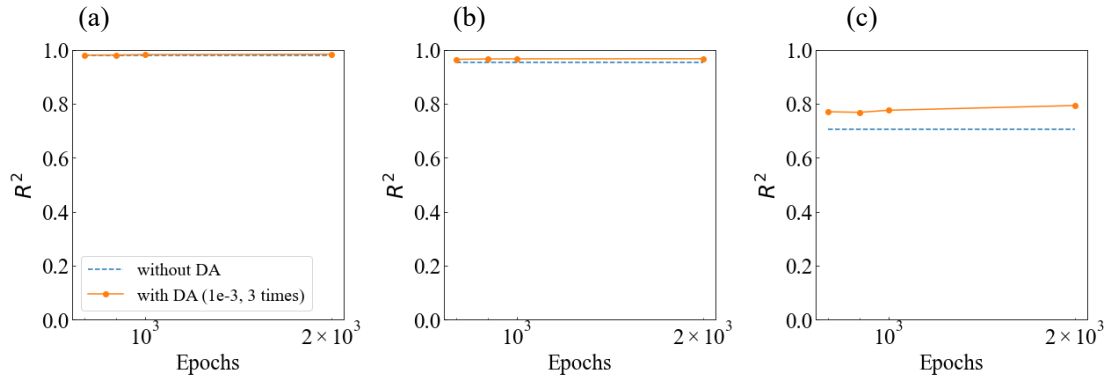


図 5.6 (a) U_{eff}^0 , (b) U_{eff}^T , (c) $1/\tau$ のテストデータにおける R^2 のエポック特性. 橙色のプロットがデータ拡張($\sigma_G^2 = 10^{-3}$, 3 倍拡張)した時の特性. 青色の破線はデータ拡張なしの時のテストデータにおける R^2 を示す.

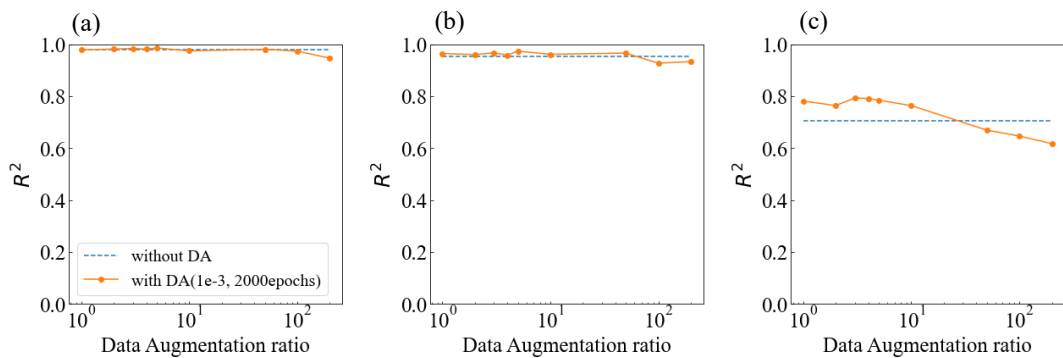


図 5.7 (a) U_{eff}^0 , (b) U_{eff}^T , (c) $1/\tau$ のテストデータにおける R^2 の n 特性. 橙色のプロットがデータ拡張($\sigma_G^2 = 10^{-3}$, 2000 エポック)した時の特性. 青色の破線はデータ拡張なしの時のテストデータにおける R^2 を示す.

5.5 ガウシアンノイズによるデータ拡張と荷重減衰による正則化の比較

5.4 ではガウシアンノイズによるデータ拡張 (DA) が予測モデルの性能向上に寄与していることを述べた. 本節ではニューラルネットワークによく用いられる典型的な正則化である荷重減衰(weight decay)による性能向上とガウシアンノイズの DA による性能向上を比較し, DA のメリット・デメリットについて考察する.

weight decay は線形回帰における L2 正則化 (ridge 回帰) と同じく重みのノルムの 2 乗に対しハイパーパラメータ λ でスケールしたものを損失関数に加える. これにより重みの行列のうち, 大きい重みの要素ほど小さくなるように重み更新が働き, 重みの極端な巨大化を防ぐことで全体的に重みの値を小さくさせ, 予測モデルの性能を向上させる. また weight decay と L2 正則化との違いは, L2 正則化では重みは入力層と出力層の間のみであるに対し, weight decay ではこれらと隠れ層との間の全ての重みを用いて計算することである.

Weight decay を用いた場合の損失関数の定義式は以下のとおりである.

$$L = C(y^{train}, y^{pred}) + \lambda \sum_i |W_i|^2 \quad (5.2)$$

ここで C は典型的な損失関数 (回帰問題ならば MSE など) であり, W は層間の重みである.

図 5.8 に DA の σ_G^2 特性と weight decay の λ 特性について示す. まず図 5.8(a)-(c)の DA と図 5.8(d)-(f)を比較して顕著に異なる点は, ハイパーパラメータ(DA では σ_G^2 , weight decay では λ)による R^2 の変化である. DA では図 5.8(a)と(d), 図 5.8(b)と(e)を比較するとどちらもハイパーパラメータが 10^{-2} では訓練・テスト両方において性能が悪化する. しかし weight decay では DA に比べて悪化が激しく, 10^{-1} では正解データと予測データがほぼ無相関となっている. これは weight decay が λ を大きくしすぎると全ての重みが 0 に近づくため正しい予測ができなかったと考えられ, 対して DA ではそのような影響は見受けられない. そのためこの比較において, DA はハイパーパラメータの変化に対しロバスト (入力に加えたノイズなどの擾乱に対して出力はほとんど影響を及ぼさない状態) であると言える. また $1/\tau$ の図 5.8(c)と(f)を比較すると U_{eff}^O と U_{eff}^T を比較した時と同じく 10^{-2} 以上で R^2 が訓練・テスト両方で悪化している. また 10^{-2} 以下の場合, それぞれの訓練・テストデータの R^2 の差分を比較すると DA の方が weight decay より小さくなっている. そのため weight decay の方がハイバリアンスな状態であり, 訓練データにオーバーフィッティングしていることがわかった. そのため DA では weight decay よりオーバーフィッティングを抑えた学習であると考えられる.

以上のことから DA は weight decay に比べてハイパーパラメータの変化に対してロバストであり, オーバーフィッティングを小さくできるメリットがあることがわかった. しか

し性能に顕著に寄与するハイパーパラメータの個数を比較すると, DA は 2 つ(σ_D^2, n), weight decay では 1 つ(λ)であり, チューニングの容易度は weight decay の方が高いため, DA ではハイパーパラメータチューニングにおいて不利であるデメリットがあり, 普遍的に DA は weight decay より優れる訳ではなく, 状況によって使い分ける必要があると考えられる。

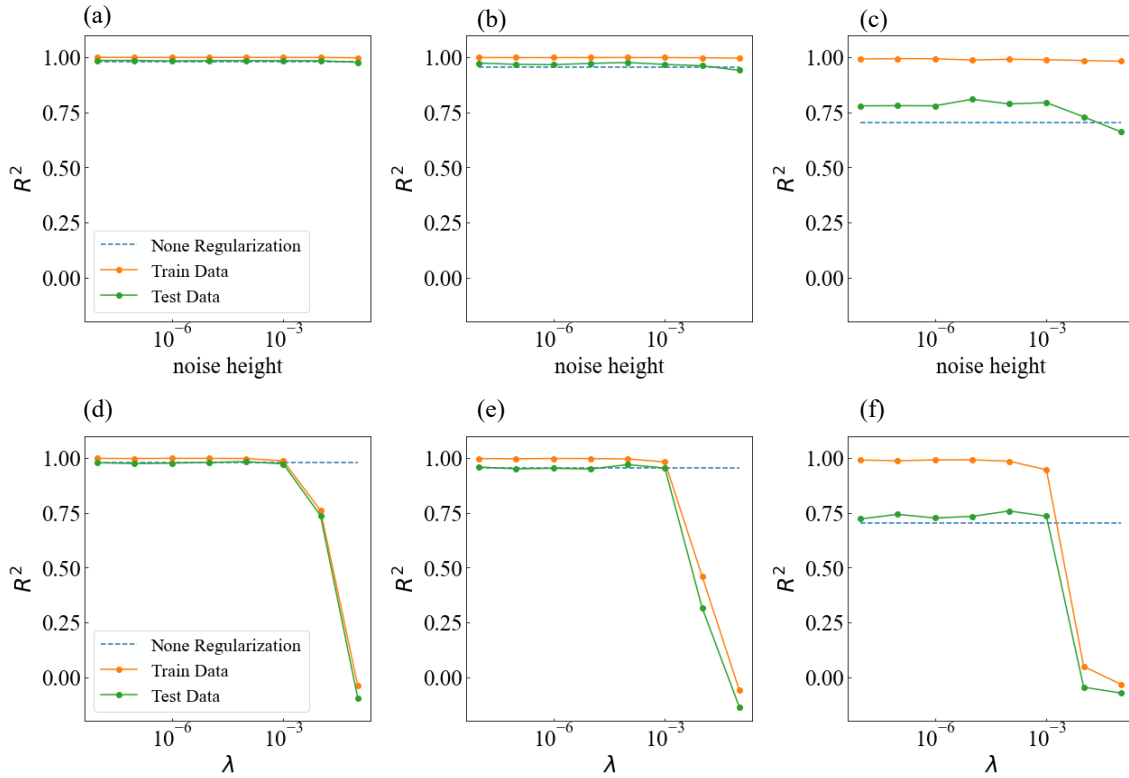


図 5.8 (a)-(c) DA による U_{eff}^O , U_{eff}^T , $1/\tau$ の R^2 . (d)-(f) weight decay による U_{eff}^O , U_{eff}^T , $1/\tau$ の R^2 . 橙色のプロットが訓練データ, 緑色のプロットがテストデータによる R^2 . 青色の破線は DA 前のテストデータの R^2 .

5.6 推定したパラメータを用いた第一原理の光学物性と実験結果の比較

5.4 にて構築した予測モデルに対し、実験で観測された誘電関数を入力したところそれぞれ $U_{\text{eff}}^0 = 4.99 \text{ eV}$, $U_{\text{eff}}^T = 7.40 \text{ eV}$, $1/\tau = 0.11 \text{ eV}$ が得られた。表 5.1 にこれらの値を他の手法で計算した結果を示した。まず U_{eff} の値は過去の報告の値の範囲(2.7~9.8 eV)の中間程度の値である。また線形応答理論に基づいて導出した例($U_{\text{eff}}^0 = 9.8 \text{ eV}$, $U_{\text{eff}}^T = 9.1 \text{ eV}$)[99]と比較すると、 U_{eff}^0 , U_{eff}^T がそれぞれ異なる値を示している点では共通しているが本研究では U_{eff}^T が U_{eff}^0 より大きく予測される点が変わった。またこれらのパラメータを用いて YIG の第一原理計算を行った結果の電子構造の比較を表 5.2 に示す。バンドギャップ $E_g = 2.71 \text{ eV}$ (実験 2.85 eV [100]), Fe イオンにおける磁気モーメント M_{FeO} と M_{FeT} はそれぞれ $4.08 \mu_B$, $4.06 \mu_B$ となり、全スピン磁気モーメントも $5.00 \mu_B$ (実験 $5.01 \mu_B$ [101], $5.26 \mu_B$ [102]) となったため、実験の電子構造を定量的に再現することに成功した。

続いて得られた電子構造から計算した誘電関数スペクトルを図 5.9 に示す。まず予測モデルに用いた実部である図 5.9(a)では第一原理計算から得たスペクトルは実験[104]を定性的に再現した。また 3.5 eV 周辺に存在しているショルダ一部も定性的に一致し、全体として特徴を捉えたスペクトルを再現できており U_{eff}^0 , U_{eff}^T は実験の誘電関数を再現するに当たって信頼度の高いものであると考えられる。しかし最大ピーク周辺での分散は実験と比べて過大評価されており、 $1/\tau$ が理想的な値に比べて大きく見積もられていると考えられる。一方虚部の図 5.9(b)では最大ピーク位置を定量的に再現するには至らなかったが、全体的なスペクトル形状と 3.1 eV 周辺のショルダ一部を定性的に再現できた。以上のことから実部・虚部ともにスペクトル形状は定性的に一致することはできたが、実部の最大ピーク周辺の分散や虚部のピーク位置の定量的な一致は実現できなかった。この原因として、まず前者は誘電関数のピーク周辺を制御する $1/\tau$ の取り扱いが不十分であったことが挙げられる。5.4 で示した DA によって $1/\tau$ の性能は向上したが、依然 U_{eff}^0 , U_{eff}^T の精度に比べて劣っており、今回の予測でも $1/\tau$ を正確に予測するには性能が不十分であったと考えられる。一方後者の原因は本研究で行なった予測モデルの訓練に虚部を用いなかったためと考えられる。2 章にて記述したとおり、実部と虚部はクラマースクローニツヒの関係で関連づけられており、実部が完全に再現できていれば虚部も完全に再現できるはずである。しかし本研究では実部が定量的に一致しなかったため、それに伴い虚部の定量一致は困難であった。従ってこの予測手法だけでは情報量不足であると考えられるため、虚部を用いた予測が必要であったと考えられる。

最後に磁気光学カー効果について第一原理計算から得られた結果を図 5.10 に示す。この結果は、スペクトル形状は実験と定性的に一致したものの最大ピーク高さは実験と比較して小さく見積もられた。一方、カー回転角にて実験ではショルダとして観測された部分 (2.5 eV 付近)では第一原理計算ではピークとして見積もられており、定量的な一致には至ら

なかった。これは磁気光学カー効果において、予測モデルの訓練に用いなかった誘電関数の非対角項が対角項に比べて支配的であったため、非対角項の実験の再現が不十分であったためと考えられる。先行研究[99]でも磁気光学カー効果の非対角項成分が支配的であると報告しているため、本研究でも同様のことが起こったと考えられる。

表 5.1 計算手法による U_{eff}^O , U_{eff}^T , $1/\tau$ の比較. それぞれの単位は eV.

基底	手法	U_{eff}^T	U_{eff}^O	$1/\tau$	Ref.
LAPW ^a	機械学習+データ拡張	7.40	4.99	0.11	This work
LCAO ^b	実験にフィッティング	2.7	2.7	-	[27]
PAW ^c	実験にフィッティング	5.7	5.7	-	[28]
LAPW ^a	線形応答理論	9.1	9.8	-	[99]

a: Linear augmented plane wave

b: linear combination of atomic orbitals

c: projector-augmented wave

表 5.2 YIG のバンドギャップ, 全スピン磁気モーメントおよび鉄原子のスピン磁気モーメントの先行計算および実験との比較. ただし鉄原子のモーメントについては, Muffin-tin 球内部の磁気モーメントであり, 格子間領域に存在する電子を含まない.

手法	E_g [eV]	M_{Fe^O} [μ_B]	M_{Fe^T} [μ_B]	M [μ_B]	Ref.
FLAPW(GGA+U)	2.71	4.08	4.06	5.00	This work
FLAPW(GGA+U)	2.77	4.27	4.16	5.00	[99]
PW(GGA+U)	-	4.12	4.02	-	[103]
実験	-	-	-	5.01	[101]
実験	-	-	-	5.26	[102]

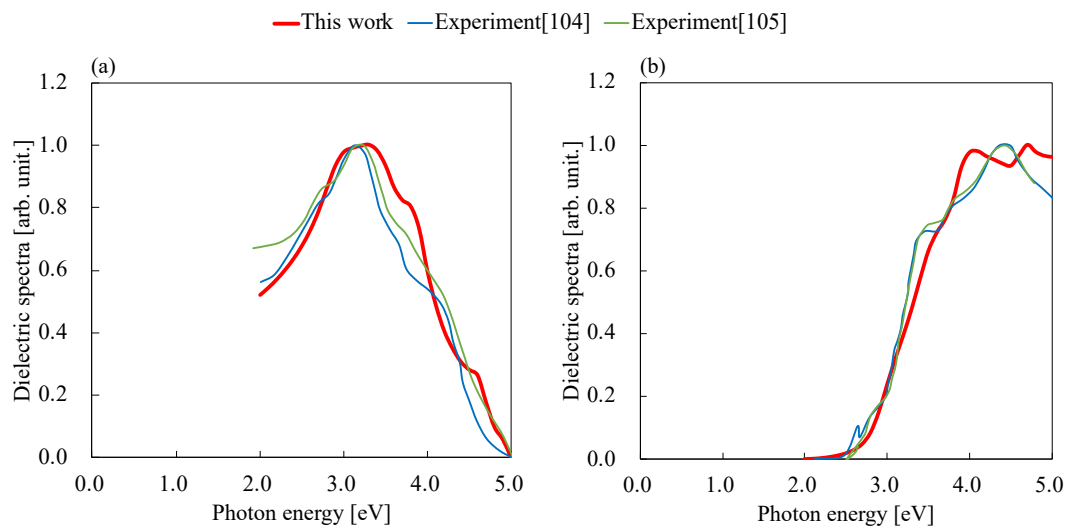


図 5.9 第一原理から計算した誘電関数 ϵ_{xx} (赤線) と実験から得られた誘電関数 (青線[104], 緑線[105]) (a) 実部(b)虚部. (a), (b)それぞれのスペクトルは 0 から 1 の範囲で規格化した.

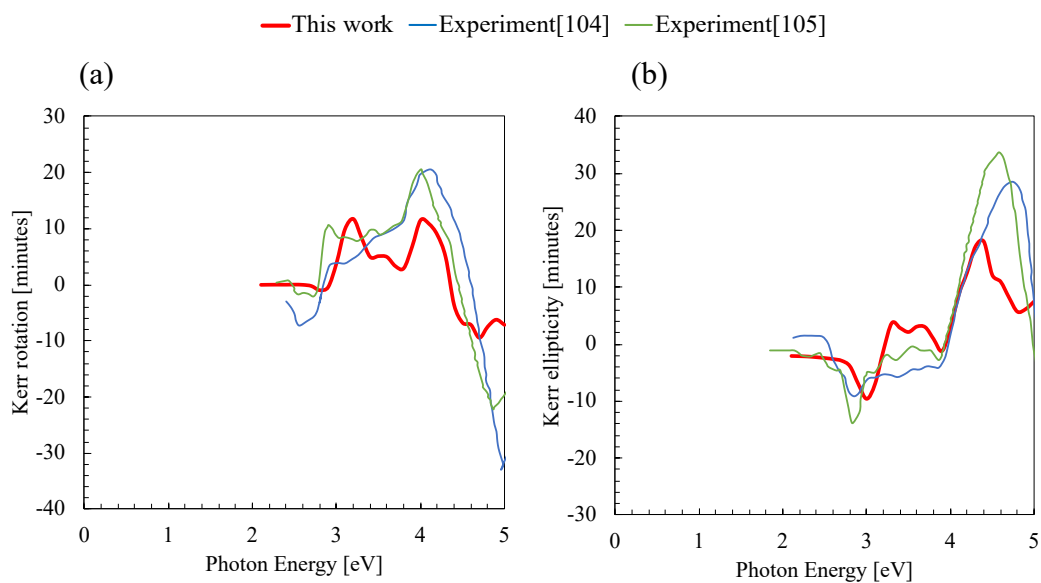


図 5.10 第一原理から計算した磁気光学カー効果のスペクトル (赤線) と実験から得られたスペクトル (青線[104], 緑線[105]) の(a) カー回転角(b)カー楕円率.

5.7 まとめ

本章では YIG に関する予測モデルの構築とデータ拡張による性能向上手法について述べた。また構築した予測モデルに実験で観測された誘電関数を入力することで U_{eff}^O , U_{eff}^T , $1/\tau$ を予測し、これらを用いた第一原理計算によって電子構造、光学特性（誘電関数、磁気光学カー効果）について実験と比較し議論した。

まず 5.2 ではランダムに選択した $(U_{\text{eff}}^O, U_{\text{eff}}^T, 1/\tau)$ の組における第一原理計算を行い、誘電関数を計算するに至るまでの計算条件について詳しく述べた。

続いて 5.3 では得られた $(U_{\text{eff}}^O, U_{\text{eff}}^T, 1/\tau)$ -誘電関数対角項 ϵ_{xx} データセットを用い、ニューラルネットワークによる予測モデルの構築を行なった。この予測モデルでは $(U_{\text{eff}}^O, U_{\text{eff}}^T)$ ではテストデータにおける R^2 で高い性能を示したが、 $1/\tau$ の R^2 では $(U_{\text{eff}}^O, U_{\text{eff}}^T)$ での R^2 にくらべて低い性能を示した。この原因の 1 つとしてデータ数不足を指摘し、データ拡張の必要性を明らかにした。

5.4 ではガウシアンノイズによるデータ拡張を本データセットに対し実施し、ガウシアンノイズの分散特性、エポック数特性、拡張倍率特性について議論した。これら 3 つの特性からテストデータにおける R^2 がもっとも高い性能を示す組み合わせを探索し、最も性能が高い組み合わせでは $1/\tau$ の R^2 が 10% 程度向上することが明らかとなった。またそれぞれの特性について極端にハイパーパラメータが大きくなると性能が悪化する理由についても議論した。

5.5 ではガウシアンノイズによるデータ拡張と **weight decay** による正則化を比較し、データ拡張のメリット・デメリットについて比較した。その結果データ拡張はハイパーパラメータについてロバストであり、 $1/\tau$ では訓練データ、テストデータによる R^2 の比較から、オーバーフィッティングを小さくしていることがわかった。これらの理由について詳しく議論した。

5.6 では 5.4 で得られた予測モデルを用い、実験で観測された誘電関数を入力することで $(U_{\text{eff}}^O, U_{\text{eff}}^T, 1/\tau)$ の最適値を予測した。またこれらを用いて第一原理計算を行い、電子構造とそれから得られる光学特性（誘電関数・磁気光学カー効果）について実験と比較しながら議論した。光学特性において第一原理計算から得られたスペクトルは実験とは定性的に一致したが、定量一致には至らなかった。この原因についてデータ選択、機械学習手法の両面から議論した。

6. 結論

YIG の理論的理解のためには MO 効果の基礎物性である誘電関数を高効率かつ高精度に理解・解明することが重要であるが、第一原理計算においてはハバードモデル (DFT+ U 法) における有効オンサイトクーロン相互作用 (U_{eff}) の値や、バンド間遷移に伴う緩和時間 ($1/\tau$) の値を実験結果に最適化できていないのが現状である。そのためこれらの変化を正確に予測できる非線形回帰が重要な鍵を握っている。そこで本研究では MI 的手法であるニューラルネットワークを応用して誘電関数から U_{eff} と $1/\tau$ を最適化する手法を開発し、この問題の解決を目指した。以下にその概略を示す。

- (1) 本手法の妥当性を確認するため、典型的な 3d 遷移金属酸化物である NiO における ($U_{\text{eff}}, 1/\tau$)-誘電関数データセットを構築し、4 層ニューラルネットワークにこれらの関係性を学習させた。この時のテストデータにおける R^2 は $U_{\text{eff}}, 1/\tau$ それぞれで 0.99 以上を示した。またこの予測モデルに実験から得られた誘電関数を入力し、 $U_{\text{eff}} = 5.52\text{eV}$, $1/\tau = 0.551\text{eV}$ を得た。この結果を用いて第一原理計算を行い得られた誘電関数は実験と定性的に一致した。また実部・虚部ともに最大ピーク位置を再現できており、実験とよく一致することから本手法の妥当性を確認した。
- (2) 複数の U_{eff} を持つ YIG に本手法を適用し、誘電関数から ($U_{\text{eff}}^O, U_{\text{eff}}^T, 1/\tau$) の最適値を予測した。 U_{eff}^O と U_{eff}^T はテストデータにおける R^2 が 0.95 以上と極めて良い予測を行なった一方、 $1/\tau$ は 0.7 程度と U_{eff} と比べて予測精度が劣った。この原因にデータ数不足を挙げデータ拡張による精度向上の必要性が明らかとなった。
- (3) ガウシアンノイズによるデータ拡張を YIG の誘電関数の訓練データに適用し、ガウシアンノイズの分散、エポック数、拡張倍率をハイパーパラメータとして最も性能が向上する組み合わせを探索した。これにより $1/\tau$ のテストデータにおける R^2 はデータ拡張以前と比べて 10% 程度の向上を確認した。またそれぞれのハイパーパラメータの特性を計算したところ、データ拡張には分散と拡張倍率の 2 つが大きな寄与をしていることがわかった。また weight decay の正則化とデータ拡張を比較したところハイパーパラメータに対し類似した振る舞いが確認できることから、データ拡張は正則化と同様の振る舞いを行い、性能向上に寄与していると考えられる。
- (4) データ拡張により得られた予測モデルに実験で観測された誘電関数を入力したところそれぞれ $U_{\text{eff}}^O = 4.99\text{ eV}$, $U_{\text{eff}}^T = 7.40\text{ eV}$, $1/\tau = 0.11\text{ eV}$ が得られた。これらをもちいて第一原理計算を行い、そこから得られた誘電関数を実験と比較したところ、スペクトル形状が定性的に一致した。また最大ピーク位置やショルダー位置などスペクトルを特徴づける部分も一致し良い予測ができていたことが示された。しかし実部において最大ピーク周りの分散が定量的に一致しないなど実験と理論の完全一致にはまだ課題があることも示された。この原因は、前者は機械学習に用いるデータに虚部を含める必要性、後者では $1/\tau$ の予測精度が不十分であることが考えられる。以上のことより今

後のさらなる発展が必要であると考えられる.

これらの結果から, 既存の線形回帰手法の拡張と計算コストを削減したパラメータ最適化手法として, 誘電関数と U_{eff} , $1/\tau$ の関係から最適パラメータを予測する機械学習手法を開発し, 誘電関数の定性的な実験の再現に成功した. しかし本研究では誘電関数対角項のみに焦点を当てていたため, 本手法によって予測されたパラメータは磁気光学カー効果など非対角項が寄与する部分などが一致しないなど依然課題が残されている. そのため光学特性とその元になる電子構造の定量的な実験の再現には, 今後の発展によって改良を重ねていく必要があると考えられる.

参考文献

- [1] Y. Wang, Richards, *et al.*, Nature Mater 14, 1026–1031 (2015).
- [2] 令和 2 年版 情報通信白書, 総務省, 2020
- [3] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B 44, 943 (1991).
- [4] S. Geller and M. A. Gilleo, J. Phys. Chem. Solids 3, 30 (1957).
- [5] A. M. Clogston, J. Appl. Phys. 31, 1985 (1960).
- [6] K. A. Wickersheim and R. A. Lefever, J. Chem. Phys. 36, 844 (1962).
- [7] D. L. Wood and J. P. Remeika, J. Appl. Phys. 37, 1232 (1966).
- [8] R. E. Macdonald and J. W. Beck, J. Appl. Phys. 40, 1429 (1969).
- [9] B. K. Larsen and R. Metselaar, J. Solid State Chem. 12, 253 (1975).
- [10] G. B. Scott and J. L. Page, Phys. Status Solidi B 79, 203 (1977).
- [11] P. Röschmann, IEEE Trans. Microwave Theory Tech. 21, 52 (1973).
- [12] Y. Murakami, T. Ohgihara, and T. Okamoto, IEEE Trans. Microwave Theory Tech. 35, 1192 (1987).
- [13] M. N. Deeter, A. H. Rose, and G. W. Day, J. Lightwave Technol. 8, 1838 (1990).
- [14] S. Higuchi, K. Ueda, F. Yahiro, Y. Nakata, H. Uetsuhara, T. Okada, and M. Maeda, IEEE Trans. Magn. 37, 2451 (2001).
- [15] Y. Shoji, T. Mizumoto, H. Yokoi, I. -W. Hsieh, and R. M. Osgood, Appl. Phys. Lett. 92, 071117 (2008).
- [16] B. J. H. Stadler and T. Mizumoto, IEEE Photonics J. 6, 1 (2014).
- [17] R. Fujikawa, A. V. Baryshev, J. Kim, H. Uchida, and M. Inoue, J. Appl. Phys. 103, 07D301 (2008).
- [18] A. Kehlberger, K. Richter, M. C. Onbasli, G. Jakob, D. H. Kim, T. Goto, C. A. Ross, G. Götz, G. Reiss, T. Kuschel, and M. Kläui, Phys. Rev. Appl. 4, 014008 (2015).
- [19] U. von Barth and L. Hedin, J. Phys. C 5, 1629 (1972).
- [20] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
- [21] K. Terakura, T. Oguchi, A. R. Williams, and J. Kuebler, Phys. Rev. B 30, 4734 (1984).
- [22] G. A. Sawatzky and J. W. Allen, Phys. Rev. Lett. 53, 2239 (1984).
- [23] O. Gunnarsson and K. Schonhammer, Handbook of the Physics and Chemistry of the Rare Earth, Elsevier, Amsterdam, Vol. 10 (1987).
- [24] J. Zaanen and G. A. Sawatzky, Can. J. Phys. 65, 1262 (1987).
- [25] A. Fujimori, F. Minami, and S. Sugano Phys. Rev. B 29, 5225 (1984).
- [26] K. C. Hass, in Solid State Physics, Academic, New York, Vol. 42 (1989).
- [27] W. Y. Ching, Z.-Q. Gu, and Y.-N. Xu, J. Appl. Phys. 89, 6883 (2001).
- [28] L.-S. Xie, G.-X. Jin, L. He, G. E. W. Böhler, J. Barker, and K. Xia, Phys. Rev. B 95, 014423 (2017).
- [29] J. Chen, H. -S. Hsu, F. -Y. Lo, Mater. Res. Express 8 026101 (2021).

- [30] M. Yu, S. Yang, C. Wu, N. Marom, npj Comput Mater 6, 180 (2020).
- [31] J. Heyd, G. E. Scuseria and J. Chem. Phys. 118, 8207 (2003)
- [32] G. Kresse, J. Hafner, Phys. Rev. B, **47**, 558, 1993.
- [33] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. Phys. Rev. B, 49:14251, 1994.
- [34] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput. Mat. Sci., 6:15, 1996.
- [35] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Phys. Rev. B, 54:11169, 1996.
- [36] M. Cococcioni and S. de Gironcoli, Phys. Rev. B 71, 035105 (2005).
- [37] E. Schrödinger, Phys. Rev. 28, 1049 (1926).
- [38] P. A. M. Dirac, Proc. R. Soc. Lond. A 117, 610 (1928).
- [39] F. Bloch, Z. Physik 52, 555 (1928).
- [40] E. Wigner and F. Seitz, Phys. Rev. 43, 804 (1933).
- [41] J. C. Slater, Phys. Rev. 51, 846 (1937).
- [42] C. Herring, Phys. Rev. 57, 1169 (1940).
- [43] J. Korricga, Physica 13, 392 (1947).
- [44] W. Kohn and N. Rostoker, Phys. Rev. 94, 1111 (1954).
- [45] J. C. Slater and G. F. Koster, Phys. Rev. 94, 1498 (1954).
- [46] J. C. Phillips and L. Kleinman, Phys. Rev. 116, 287 (1959).
- [47] P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
- [48] W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
- [49] E. Wimmer, H. Krakbohrer, M. Weinert, and A. J. Freeman, Phys. Rev. B 24, 864 (1981).
- [50] M. Weinert, E. Wimmer, and A. J. Freeman, Phys. Rev. B 26, 4571 (1982).
- [51] A. D. Becke, J. Chem. Phys. 98, 1372 (1993).
- [52] A. Georges, G. Kotliar, W. Krböhrth, and M. Rozenberg, Rev. Mod. Phys. 68, 13 (1996).
- [53] O. K. Andersen, Phys. Rev. 12 3060 (1975).
- [54] M. Weinert, J. Math. Phys. 22, 2433 (1981).
- [55] J. D. Jackson, Classical electrodynamics, Wiley, New York, 3rd edition (1999).
- [56] J. Hubbard, Proc. R. Soc. Lond. A 276, 238 (1963).
- [57] J. Hubbard, Proc. R. Soc. Lond. A 277, 237 (1964).
- [58] J. Hubbard, Proc. R. Soc. Lond. A 281, 401 (1964).
- [59] J. Hubbard, Proc. R. Soc. Lond. A 285, 542 (1965).
- [60] J. Hubbard, Proc. R. Soc. Lond. A 296, 82 (1967).
- [61] J. Hubbard, Proc. R. Soc. Lond. A 296, 100 (1967).
- [62] P. W. Anderson, Phys. Rev. 124, 41 (1961).
- [63] S. Alexander and P. W. Anderson, Phys. Rev. 133, A1594 (1964).

- [64] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B 44, 943 (1991).
- [65] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. 61, 689 (1989).
- [66] J. L. Erskine and E. A. Stern, Phys. Rev. B 12, 5016 (1975).
- [67] W. Reim, J. Magn. Magn. Mater. 58, 1 (1986).
- [68] R. Kubo, J. Phys. Soc. Jpn. 12, 570 (1957).
- [69] P. M. Oppeneer, T. Møhrer, J. Sticht, and J. Kübler, Phys. Rev. B 45, 10924 (1992).
- [70] Rosenblatt, F. Psychological Review, 65(6), 386-408 (1958).
- [71] G., B., & P., B. Amphibia-Reptilia, 5(3), 225-239 (1984).
- [72] H. D. Block, B. W. Knight, Jr., and F. Rosenblatt, Rev. Mod. Phys. 34, 135 (1962)
- [73] Rumelhart, D., Hinton, G. & Williams, R. Nature 323, 533–536 (1986).
- [74] G. E. Hinton, R. R. Salakhutdinov, Science, 5786, 313, 504-507 (2006).
- [75] B. Le Cun, *et al.*, Proceedings of the Advances in Neural Information Processing Systems (NIPS), pp. 396–404 (1989).
- [76] Andrzej Maćkiewicz, Waldemar Ratajczak, Computers & Geosciences, 19, 3, 303-342. (1993)
- [77] Deng L. IEEE Signal Processing Magazine. 29(6):141–2 (2012).
- [78] Alex Krizhevsky, Learning multiple layers of features from tiny images. (2009)
- [79] C. Shorten, T. M. Khoshgoftaar, Big Data 6, 60 (2019).
- [80] A. Mikołajczyk and M. Grochowski, 2018 International Interdisciplinary PhD Workshop (IIPhDW), 117-122. (2018)
- [81] S. S. Yadav, S. M. Jadhav, J Big Data 6, 113 (2019).
- [82] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 702-703 (2020).
- [83] M. Arslan, M. Guzel, M. Demirci, & S. Ozdemir, 2019 4th International Conference on Computer Science and Engineering (UBMK) (2019).
- [84] V.Nair, G. E. Hinton, Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010.
- [85] M. Matsumoto, T. Nishimura, ACM Transactions on Modeling and Computer Simulations 8 3–30 (1998).
- [86] R. Herbert, and S. Monro, The Annals of Mathematical Statistics 22, no. 3, 400–407 (1951).
- [87] P. Kingma and J. Ba, CoRR abs/1412.6980 (2015).
- [88] J. Snoek, H. Larochelle, and R. P. Adams, Proceedings of the 25th International Conference on Neural Information, NY, USA, 2951–2959 (2012).
- [89] P. Vidnerová, R. Neruda, H. Yang, K. Pasupa, A. C. S. Leung, J. T. Kwok, J. H. Chan, I. King, Neural Information Processing. ICONIP 2020 Lecture Notes in Computer Science, vol 12534. Springer, Cham (2020).
- [90] Kvalseth, T. O., The American Statistician, 39(4), 279–285 (1985).
- [91] S. L. Dudarev, A. I. Liechtenstein, M. R. Castell, G. A. D. Briggs, and A. P. Sutton, Phys. Rev. B,

- 56, 4900 (1997).
- [92] R. J. Powell and W. E. Spicer, Phys. Rev. B 2, 2182 (1970).
 - [93] V. I. Anisimov, J. Zaanen, and O. K. Andersen. Phys. Rev. B, 44, 943 (1991).
 - [94] W. E. Pickett, S. C. Erwin, and E. C. Ethridge. Phys. Rev. B, 58, 1201 (1998).
 - [95] L. Wang, T. Maxisch, and G. Ceder. Phys. Rev. B, 73, 195107 (2006)
 - [96] A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson, and G. Ceder, Phys. Rev. B, 84, 045115 (2011).
 - [97] B.-C. Shih, T. A. Abtew, X. Yuan, W. Zhang, and P. Zhang. Phys. Rev. B, 86, 165124 (2012).
 - [98] S. Geller and M. A. Gilleo, J. Phys. Chem. Solids 3, 30 (1957).
 - [99] H. Nakashima, A. -M. Pradipto, T. Akiyama, T. Ito, and K. Nakamura, AIP Advances 10, 045029 (2020).
 - [100] R. Metselaar and P. K. Larsen, Solid State Commun. 15, 291 (1974).
 - [101] E. E. Anderson, Phys. Rev. 134, A1581 (1964).
 - [102] S. R. Nimbore, D.R. Shengule, S. J. Shukla, G. K. Bichile, and K. M. Jadhav, J. Mater. Sci. 41, 6460 (2006).
 - [103] L.-S. Xie, G.-X. Jin, L. He, G. E. W. Bauer, J. Barker, and K. Xia, Phys. Rev. B 95, 014423 (2017)
 - [104] F. J. Kahn, P. S. Pershan, and J. P. Remeika, Phys. Rev. **186**, 891 (1969).
 - [105] S. Wittekoek, T. J. A. Popma, J. M. Robertson, and P. F. Bongers, Phys. Rev. B 12, 2777 (1975).

謝辞

本研究を行うにあたり, 終始多大なるご指導, ご鞭撻を賜りました三重大学大学院工学研究科物理工学専攻 中村浩次教授ならびに名和憲嗣助教, 三重大学大学院教育学研究科情報教育課程 萩原克幸教授に, 心より深く感謝の意を申し上げます. 中村教授, 名和助教には, コロナ禍で先行きが見えない中, 私の研究の進捗状況に関して助言をくださりましたこと, 国内学会や国際会議など多大なご支援と経験の機会を与えてくださったこと重ねて感謝いたします. また萩原教授には, 物質・量子・計測データ科学セミナーにて機械学習, アルゴリズムに関して議論・助言を下さりましたこと感謝いたします.

また, 本修士論文の副査を引き受けてくださった三重大学大学院工学研究科物理工学専攻 佐野和博教授には謹んで感謝いたします.

最後に研究室での生活においてお世話になりましたナノデザイン研究室の大学院生, 学部生の皆様に深く感謝いたします.