

平成 30 年度 修士論文

添削者を困らせることのない
修士論文の書き方の研究

名古屋大学大学院理学研究科
素粒子宇宙物理学専攻宇宙地球物理系
宇宙線物理学研究室

博士課程（前期課程）2 年
学籍番号 123456

奥村 暁

2019 年 1 月 8 日

概要

ここには論文の概要（abstract）を書きます。論文の先頭なので早い時期に書き始める人がいますが、論文の結論や論理展開はなかなか執筆終盤まで固まりません。そのため、論文の流れや結論がかなり明確になった最終段階で書くようにしましょう。

概要は論文全体の内容を短文で説明するものですので、研究の背景と目的、研究内容、結果と結論などが全て網羅されている必要があります。ここを読んだだけで、論文の中身が大雑把に把握できるようにすることが大切です。原則として改行せずに 1 段落で書きますが、これは複数段落に分けて書くような文章を無理やり 1 段落に合体させるということではありません。1 段落で流れるように書いてください。

目次

图目次

表目次

第 1 章

序論

英語で言うところのイントロダクションです。通常、「序論」(introduction)で始める場合は「結論」(conclusion)という章で締めます。もし「はじめに」で始まる場合は「おわりに」です。

この章では研究の背景や課題などを簡潔に説明します。2 から 4 ページもあれば十分ですし、細かく節に分ける必要もありません。この章で必要なことは、なぜこの論文が書かれたのか、過去の研究に対する位置付け・課題は何か、この研究でどこまでを明らかにしようとしているのかを少ないページ数で説明することです。

このような序論の存在しない修士論文はたくさん存在しますが、何十ページにもなる修士論文では研究の位置付けや課題がどこに書かれているのか読者は見失いやすくなります。先頭に独立した章で簡潔に道筋を示すことで、続く章を読者が読みやすくなります。

第 2 章

深層学習による病理画像の診断支援

病理画像をデジタルで保存することが始まったのは数十年前になる。これによって遠隔地でも診断することができるようになったり、情報を共有することができるようになり、複数の医師で診断しミスを防止するセカンド・オピニオンが容易になった。計算機科学の分野の側面ではデータを収集することができるようになり、研究が盛んに行われることになった。その後は、様々な病理データでより改良されたアルゴリズムの提案が行われている。

細胞組織の形態を観察するための病理染色ではヘマトキシリン・エオジン染色 (HE 染色) が一般的に用いられる。細胞核を青紫色に染色し、細胞質をピンク色に染色する。正常から異常に変化していくと、細胞核が過度に増殖したり、細胞質の形が崩れたりすることで、その特徴を機械学習によって精度よく検出するための研究が行われている。

これまでは、核の形やテクスチャーからパターンマッチングなどの画像処理によって腫瘍を検出する研究されてきたが、近年になって画像処理に大きなブレイクスルーが起きたことをきっかけに、新しい手法で解析するようになってきた。そのブレイクスルーがディープラーニングである。

2.1 ニューラルネットワーク

人間の脳にはニューロンと呼ばれる神経細胞が 1000 億個以上あり、それぞれが複数のニューロンと電気信号で情報を伝達している。また脳には電気信号を受け渡すシナプスという場所がある。ニューロンとシナプスで行われる演算を模倣したアルゴリズムを作ることができれば人間のような思考や認識をコンピュータを使って再現できると考えた。そのアルゴリズムがニューラルネットワークである。

2.1.1 多層パーセプトロン

ニューラルネットワークは入力層、出力層、隠れ層から構成され、層と層の間にはニューロン同士のつながりの強さを示す重みがある。非線形問題を扱うために 1986 年 Rumelhart によって考案されたのが、パーセプトロンを複数つなぎ合わせ入力と出力以外に隠れた層を持つ多層パーセプトロン (Multi-layer perceptron: MLP) である。

ニューラルネットワークで多層パーセプトロンの層を全結合 (fully connected: FC) 層とも呼ぶ。Figure 3.1 における丸や矢印はそれぞれノード (またはニューロン) と重み (または結合) と呼び、ともに数値である。例えば画像を分類しようと思えば、各ピクセルの画素数を各ノードに入力する (28×28pix のグレースケール画像であれば 784 個のノードが必要)。データが入力層に入ってくると、その値に重みをかけ、活性化関数と呼ばれる関数を通し結果を出力する。これを繰り返し出力層に書き出す。各層の重みの値によって出力結果は異なってくる。出力層のノードは区別したいクラス数分用意し、各ノードの出力値が各クラスに属している確率を表す。学習については誤差逆伝播法を利用する。ニューラルネットワークの出力値と正解データとの比較をした時に、どれだけ正解から離れているかを評価する損失関数 (Loss) を使って、損失関数が小さくなるようにノード間の重みを勾配降下法によって更新

する。

2.1.2 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) は、1998 年には既に LeNet と呼ばれるネットワークで実装されていた [3]。従来の画像認識では、画像から特徴を抽出し、それをニューラルネットワークにかけるいわゆる特徴量設計が必要で、ここをいかにうまく設計するかがポイントであったが、CNN は特徴量設計から識別までを end-to-end で行うことができることが最大の強みである。人間が物体を認識することをコンピュータにも計算させるには、画像の特徴的な部分を切り分けて数値化させる必要がある。例えば、カラー画像の場合、RGB の 3 色 (3 チャンネル) を組み合わせた画像で認識をしている。このようなフィルターの畳み込み計算を行うと、フィルターごとに異なった画像の特徴を抽出して数値化する。これが畳み込み (convolution) である。その後、画像のサイズを小さくしてコンピュータが計算コストを減らし、微小な変化に対してロバストになる仕組みしてプーリングという方法を用いる。学習には MLP と同様に誤差逆伝播法を用いる。

2.1.3 再帰的ニューラルネットワーク

再帰的ニューラルネットワーク (Recurrent Neural Network: RNN) は時系列解析や自然言語処理に利用されるニューラルネットワークである。内部にループを持つことで過去の情報を保持しておくことができる。時系列の入力 $x = (x_1, \dots, x_T)$ があつた時に、出力 $y = (y_1, \dots, y_T)$ と隠れ層のベクトル $h = (h_1, \dots, h_T)$ をそれぞれ以下の式で計算する。

$$h_t = H(W_{ih}x_t + W_{hh}h_{t-1} + b_h) \quad (2.1.1)$$

$$y_t = W_{ho}h_t + b_o \quad (2.1.2)$$

W は重みの行列であり (W_{ih} は入力と隠れ層間の重み行列), b はバイアス項である。そして H が活性化関数である。

RNN は過去の情報をどこまでさかのぼって関連性を見つけるかを判断することができないため、時系列データが長くなるほど、その長期の依存性を学習するには人が慎重にパラメータを設計する必要があるなど、学習が難しくなるという問題があつた。この問題を解決するために Long-short term memory (LSTM) が提案された。LSTM も RNN の一種ではあるため、繰り返し構造を持ち、3 つのゲートを持つ層からなっている。

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2.1.3)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2.1.4)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (2.1.5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (2.1.6)$$

$$h_t = o_t \tanh(c_t) \quad (2.1.7)$$

ここで σ はシグモイド関数であり、次の式で定義される。

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.1.8)$$

f_t で示される層は忘却ゲート層と呼ばれ、過去の情報で捨てるべき情報を判断する。これはシグモイド層によって行われ、0 と 1 の間の値を出力し、0 は完全に忘れる。1 は完全に維持するという意味である。 i_t や c_t で示される層は、入力ゲート層と呼ばれ、 i_t で新たに入力された情報から、どの情報を更新するかを判断し c_t で古い情報を

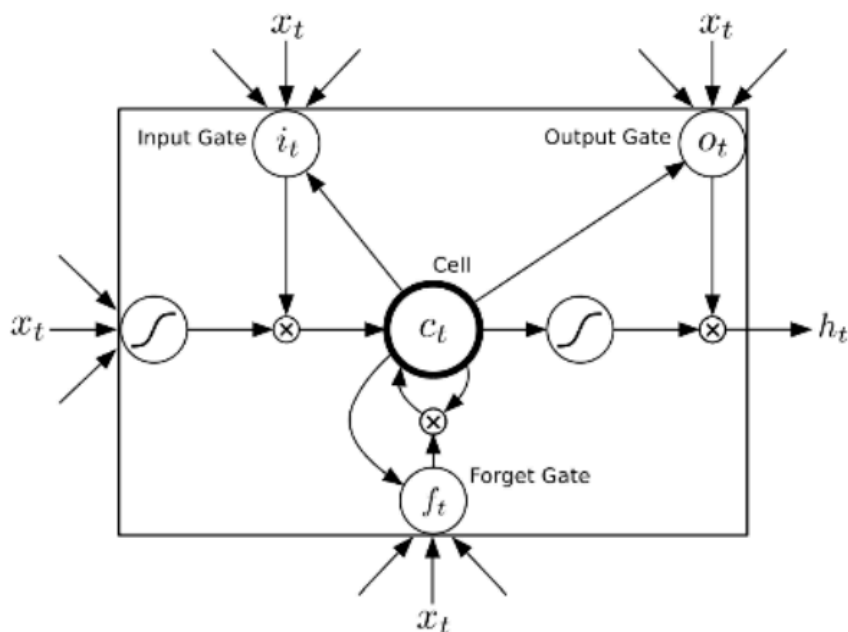


Fig. 2.1.1

落とし新しい情報を加え、値を更新する。最後が o_t や h_t で示される層で、出力ゲート層と呼ばれる。まず何を出すべきかを o_t のゲートで判断して、 c_t に \tanh を適用して掛けることで出力が計算される。

2.2 画像認識におけるディープラーニング

Deep Learning とは Deep Neural Network(DNN) を指すことが多い。この"Deep"とは、ニューラルネットワークの層が深いことに由来している。Figure 3.3 に画像認識タスクの精度の近年の推移を示す。これは ImageNet Large Scale Visual Recognition Challenge (ILSVRC) と呼ばれる世界的な画像認識のコンペティションである (2010 年から始まった)。カテゴリ数は 1000 クラスで、画像枚数は 120 万枚の訓練データと 15 万枚のテストデータが用意されている。2011 年と 2012 年は約 10% もの大差で AlexNet[4] が優勝している。これがディープラーニングの始まりである。AlexNet は 5 つの畳み込み層と 3 つの全結合層を持っている。2014 年には VGGNet[5] や GoogLeNet[6] が 9 割の精度を超えた。VGGNet は AlexNet(8 層) よりさらに深い構造 (19 層) であり、GoogLeNet は 22 層もある。そして 2015 年には ResNet[7] が人間の精度をも超える認識精度を達成した。ResNet は GoogLeNet よりもさらに深く 152 層もある。一般に、層を深くすることは簡単ではなく、勾配消失や過学習といった問題が起こる。CNN を複数回かけて検出を行う場合、CNN の浅い側では空間分解能はあるが抽象的な情報が少ない。深い側では意味論的な情報は取得できる (ポーズ、変形など) が空間分解能が小さいため幾何学的な情報が失われる。アーキテクチャの進化の方向は①層が深くなること ② FC 層の使用を避けること又は Inception モジュールの使用によってパラメータ数を削減すること ③ ResNet などのショートカット接続によって学習効率をあげること事前学習・転移学習を行うことでモデルの精度を上げる取り組みがある。当然ながらこの場合事前学習のデータセットと適用データとの間に類似性があると良い。

画像処理におけるディープラーニングでは大きく 3 つのタスクがあり、それぞれ、クラス分類、物体検出、セグメンテーションである。以下に詳細を述べる。

クラス分類

カテゴリ分けをすること

物体検出

物体検出とは **Bounding Box** で物体の位置とその物体の種類を特定する方法である。歴史的には幾何的情報、手動特徴量、そしてそのカスケードを利用していた。その後、HOG や SIFT など局所特徴量を抽出する方法を設計するようになったが、これは深い専門知識を必要とした。また広い範囲でオブジェクトを正確に検出する方法は、メモリ容量と処理時間に課題がある。現在は **Deep Neural Network** になりデータのみから抽象的な特徴量を複数得ることができる。一般物体認識の場合は数千のカテゴリを学習して **Top Error Rate** が 2% 以下と人間よりも認識精度が高いが、物体検出においては、今はカテゴリがせいぜい数百程度くらいまででも認識精度が人間よりも低くなってしまう。また物体検出は精度を上げるために処理に時間がかかるアルゴリズムであることが多いため、リアルタイムに物体検出を行う時は、速度と精度でトレードオフが生じてしまう。

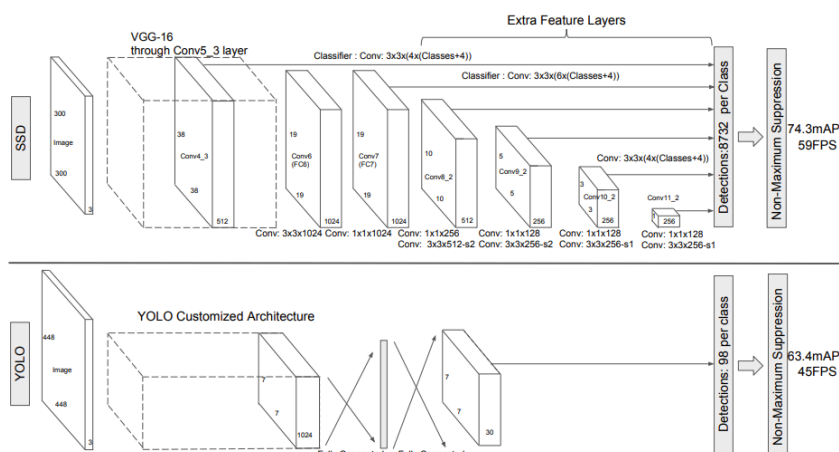


Fig. 2.2.1

セグメンテーション

セマンティックセグメンテーションとは、画像を画素レベルで認識することである。画像内の各画素をオブジェクトクラスに割り当てる手法である。セマンティックセグメンテーションの手法についてディープラーニング以前では、Texton Forests や Random Forests に基づいた分類を行っていたが、物体検出と同様に CNN が登場してからは、高精度なセグメンテーションが実現するようになった。CNN を使ったセグメンテーションの手法で一般的に使われるようになったものが Unet である (Figure ??)。この Unet は文字通り U の形をしたネットワークであることが特徴で、2つのアーキテクチャーからできている。一つ目がエンコーダーのアーキテクチャーで CNN とプーリングで特徴を抽出しながら次元を削減していき、2つ目のデコーダーのアーキテクチャーで画像をセグメンテーションの結果になるように復元する。ここで問題になることが、プーリングをすることで位置情報を消してしまっているので、この位置情報を利用して画像を復元するためには、エンコーダーとデコーダーで画像サイズが同じところ同士をショートカットで接続することが Unet 構造の優れている点である。

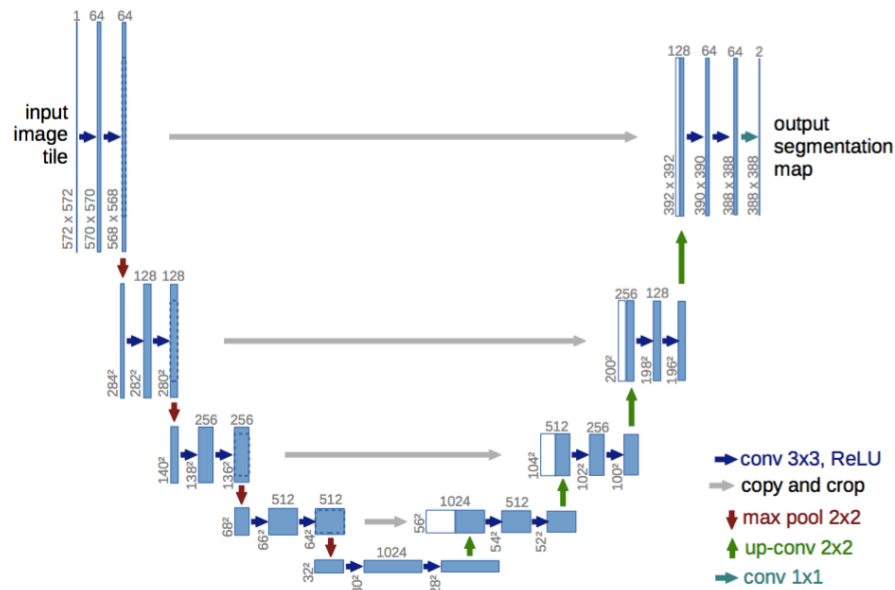


Fig. 2.2.2

2.3 深層学習による 3 次元画像解析

ディープラーニングを医療画像に応用するコンペティションが世界で行われているが、その半数が 3D 医療画像の解析になっているほど需要が高まっている。その理由は、現在解決しなくてはならない課題があるからである。まずは 2 次元画像と違って、処理すべきデータが大きいということである。そのため学習するパラメータをなるべく少なくする工夫がされている。また 3 次元画像には、動画または、ボリューム画像があるが、2 次元画像とその深さ方向（動画であれば、時間方向）には異方性があることから、機械学習の方法に工夫が必要になる。今まで考案されている手法として、2DCNN を拡張した 3DCNN、また CNN と時系列解析でよく用いられる LSTM を組み合わせた手法と、LSTM 内部に CNN を組み込んだ手法、それらをすべて組み合わせた手法が考案されている。LSTM の研究も盛んに行われているため、その改良モデルが数多く存在する。特に、LSTM の学習効率を上げた GRU(Gated Linear Unit) や、順方向だけでなく逆方向の時系列も計算に入れる BiLSTM が時系列解析の精度向上になっている報告がある。

2.3.1 3DCNN と Stacked Convolution

2 次元画像が深さ方向に連続している 3 次元画像の特徴を抽出するために、2 次元の CNN を拡張して、3 次元のカーネルを使って畳み込みを行う、3DCNN を利用した手法がある。

2.3.2 LSTM と 2DCNN の組み合わせ

時系列解析に使われる LSTM を用いて 3 次元の画像を解析することができる。これはよく動画の解析で行われることがある。つまりフレームごとの画像の特徴を 2 次元の CNN で計算してから時系列情報を LSTM で解析することで、画像の時系列解析を行うことができる。これを 3 次元の医療画像で CT や MRI で適用する研究も行われている。3DCNN のデメリットであったパラメータの増大を 2DCNN と LSTM の組み合わせで解決することが

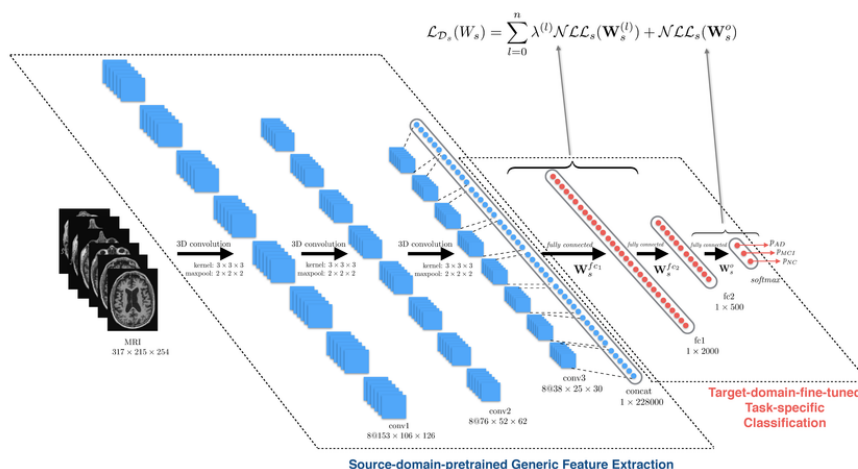


Fig. 2.3.1

できる。

2.4 教師なし学習

機械学習の手法には、上記で説明したように、ラベルの貼られているデータセットを用いて学習することを教師あり学習と呼び、その反対で、データセットはあっても、そのデータセットの特性を示したラベルが与えられていない場合のデータセットを用いて学習することを教師なし学習と呼ぶ。

2.4.1 Autoencoder

教師なし学習で画像の特徴を抽出する方法としてオートエンコーダがある。画像の場合におけるオートエンコーダの手法とは、ある画像から情報を圧縮する「エンコーダ」と言われる部分と、その圧縮した情報から画像を復元する「デコーダ」の二つからなる。入力とデコーダから復元された画像が同じ画像になるようにニューラルネットワークで学習させる。この学習の結果、潜在変数は似てる画像どうして近い値になるように変化し、この分布を見れば画像の分類を教師ラベルがなくても、学習を行うことができる。

2.4.2 Variational Autoencoder

本研究では、このオートエンコーダの派生である。Variational Autoencoder(VAE)を利用した。これはオートエンコーダの「エンコーダ」と「デコーダ」は同じネットワーク構造であるが、データセットの潜在変数の分布が、正規分布になるような制約を加えて学習を行う手法である。こうすることで Autoencoder の潜在変数では分布の距離に意味がないが、VAE では正規分布に埋め込まれるため、画像の類似度を分布が表現することができる場所が特徴である。

2.4.3 敵対的生成ネットワーク

敵対的生成ネットワーク (Generative Adversarial Network: GAN) は 2014 年に提案された手法である。概念図のように GAN では Generator と Discriminator の 2 つのネットワークがある。Generator は訓練データと同じような画像を生成するネットワークで Discriminator は、入力されたデータが訓練データから来たものか Generator

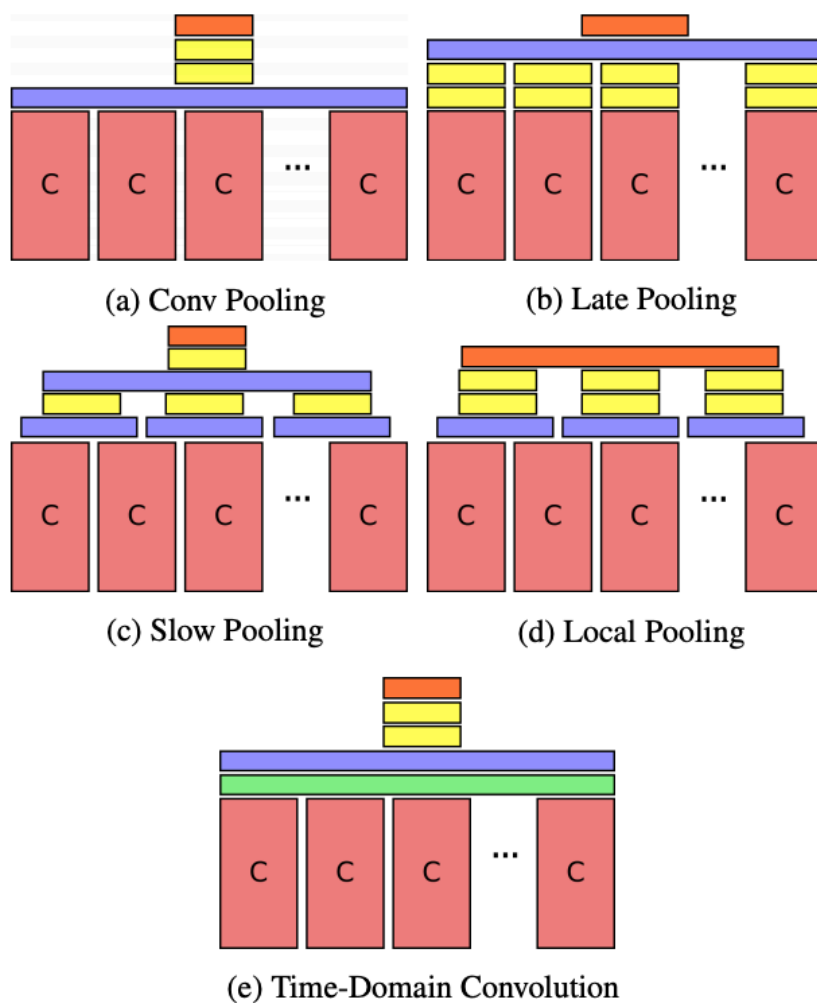


Fig. 2.3.2

で生成されたものかを識別するように学習する。VAE よりも GAN の方が細部まで鮮明に画像を生成することができる。しかし GAN は計算時間がかかるという問題や、Discriminator か Generator のどちらかが強くなってしまいうなど、学習が安定しない問題があるため、これについての多くの研究報告がされている。

2.5 半教師あり学習

弱教師あり学習とも呼ばれる。これは、教師あり学習と教師なし学習を組み合わせる学習方法である。こうすることでデータに教師ラベルをつけているものが少数であっても、データの特徴を学習しながら少量のラベルで識別境界を決めることができる。GAN や VAE の考え方を発展させてネットワークを構築することが考えられる。

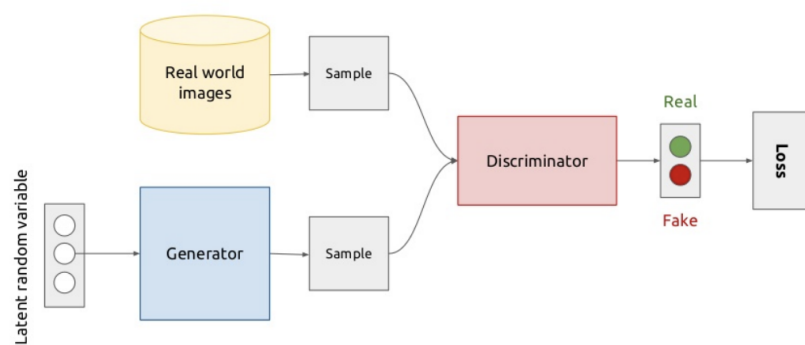


Fig. 2.4.1

第 3 章

自分の研究本体を述べるところ

ここは自分のやった研究を述べる章です。実際の中身に合わせて章を複数立てにする場合もあると思います。「議論」の章を別に分ける場合は、この章では得られた結果までを記述し、その結果に対する議論は「議論」の章に回すのが良いでしょう。この章は必ずしも 1 つの章のみである必要はありません。研究内容に応じて、複数の章に分割することも一般的に行われます。

修士論文で大切なことは、第 ?? 章や第 ?? 章で述べた伏線（研究の目的と動機）を回収するべく、きちんと研究内容を順序立てて書き、また自分の貢献を明確にすることです。論文全体で論理展開がきちんとしていれば良いので、必ずしも実際に行った実験などの時系列でこの章を書き進める必要はありません。また修士論文としての完成度が大切ですので、修士論文のテーマに直接関係のない自分のやったことを無理に混ぜる必要もありません。

第 4 章

L^AT_EX の使い方

本章では、L^AT_EX の使い方を以下説明します。ここでの表示例は本 PDF を読むだけではどのような L^AT_EX コードに対応しているか分かりませんので、**main.tex** や **LaTeX.tex** の中身を参照してください。

この PDF 文書中に **command** のような書体で記載されているものは、L^AT_EX ソース中で実際に入力するコマンドやファイル名を示しています。

4.1 節の使い方

`\section` や `\subsection` を使うと「節」(section) と呼ばれる構造を作ることができます。長い章を分割して論理展開を分かりやすくする目的で使います。

文中で節を参照するときは、**section** であっても **subsection** であっても「節」と呼び、「?? 節」や「第?? 節」のように書きます (`ref` コマンドの使用は次節参照)。章を参照するときは「?? 章」や「第?? 章」とします。

4.2 図の使い方

論文中に図を入れるときは、**figure** 環境を使用します。画像形式は図 ?? のような JPEG (主に写真などに最適) や PNG (色数の少ない画像に最適) に加え、図 ?? のように PDF (グラフなどに最適) も使うことができます。実際の使い方は、この L^AT_EX のコードを読んでください。**EPS** 形式はいまどき誰も使いません。古い L^AT_EX の本や年寄りに騙されないでください。

図を文中で参照したいときは `ref` コマンドを使用して、「図 ??」のようにすることができます。この部分は L^AT_EX 中で実際には `図~\ref{fig_CTA}` と書いています。「図」と `ref` の間に ~ を入れるのは、「図」と図番号の間で改行を防ぐためです*¹。

figure 環境で図を挿入する場所は、初めてその図を言及する段落の直後、もしくは直前です。あまりに離れた場所に図を挿入すると読者はどこに図があるかを探さなくてはならず、読むのが困難になるからです。

場合によっては複数の図を並べたいこともあるでしょう。そのようなときは、**subfigure** 環境を使って図 ?? のようにすることができます。**minipage** 環境でも似たようなことができますが、**subfigure** を使うと小番号を自動で付与したり、「図 ??」のように、小番号を参照することができます。

またせっかく図の並べ方が分かったので、同じ図を PDF、PNG、JPEG にして図 ?? にて比較してみましょう。それぞれの画像の特徴が分かります。また図 ?? は参考のため **subfigure** ではなく **minipage** 環境を使って作っています。

*¹ このようにチルダを入れる手法は、人名の姓名の間や数値と単位の間で改行を防ぐのにも広く使われます。

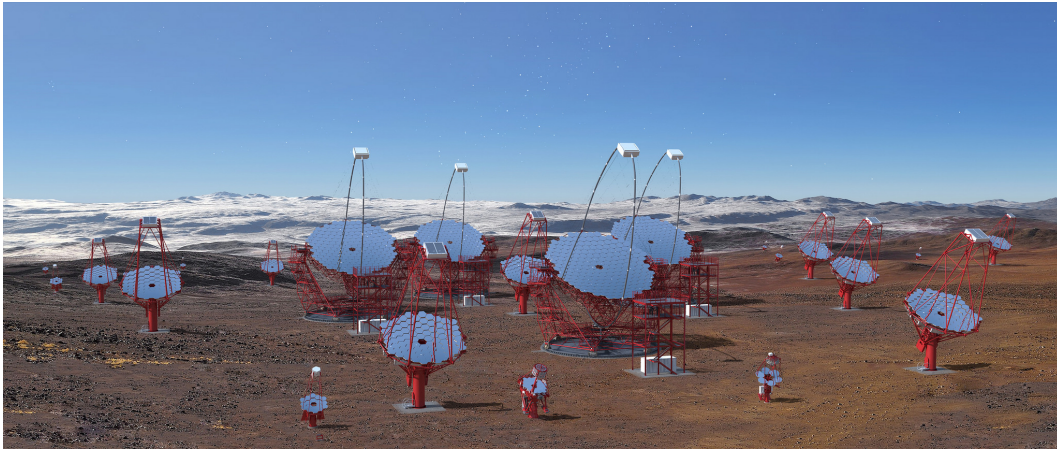


Fig. 4.2.1 CTA の完成想像図 (画像提供 : G. Pérez、IAC、SMM)。JPEG (ビットマップ画像) なので、出力 PDF で拡大するとドットが見えます。

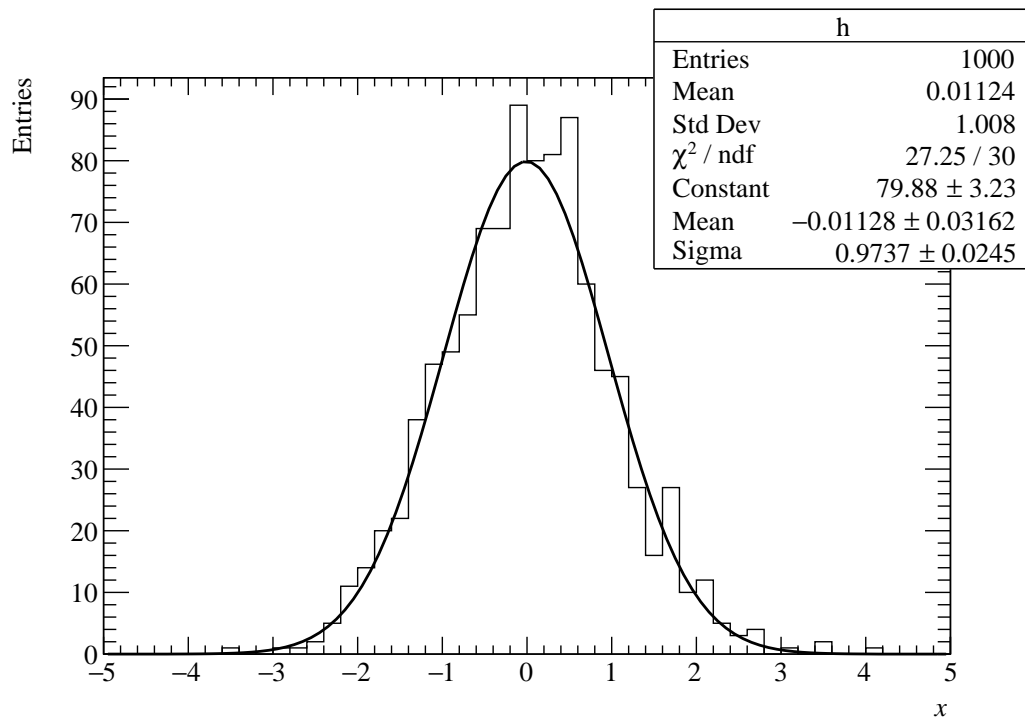


Fig. 4.2.2 ガウシアンでヒストグラムをフィットした例。PDF (ベクター画像) なので、出力 PDF で拡大しても滑らかです。また文字列も PDF 中で検索することができます。

4.3 表の使い方

表 ?? に、L^AT_EX でどのように表を作成するかの例を示します。実際にどういう L^AT_EX コードがこの表に対応するのかは、ファイルの中身を眺めてください。

論文で使う表の一般的な注意点として、あまり罫線をたくさん使いすぎないことです。日本では全てのセルの周辺に罫線を使う傾向があり、最悪、表 ?? のようになります。窮屈になるので、このような罫線の多用はやめましょう。

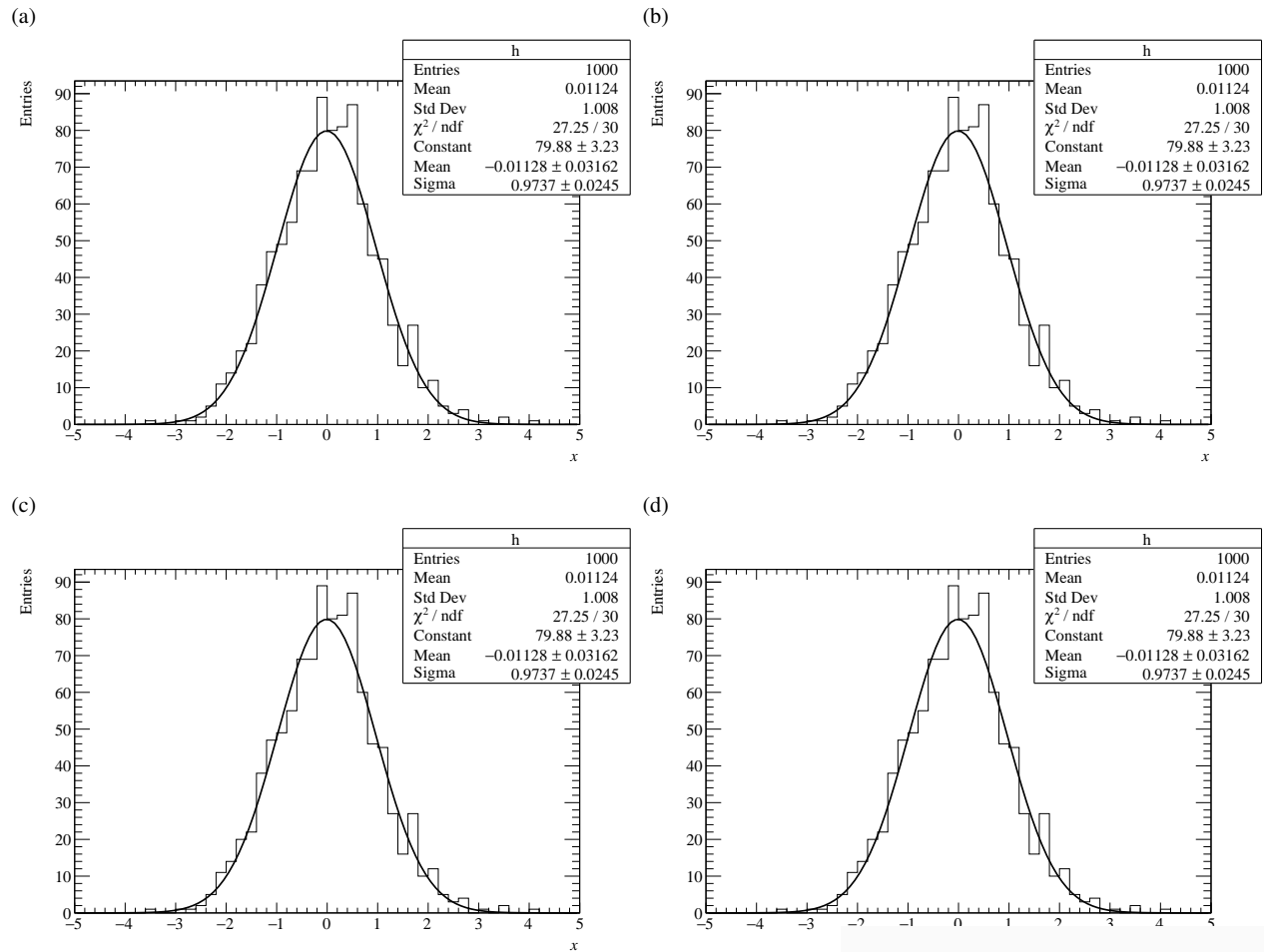


Fig. 4.2.3 複数の図を並べた例。(a) ガウシアンフィット。(b) 同じもの。(c)

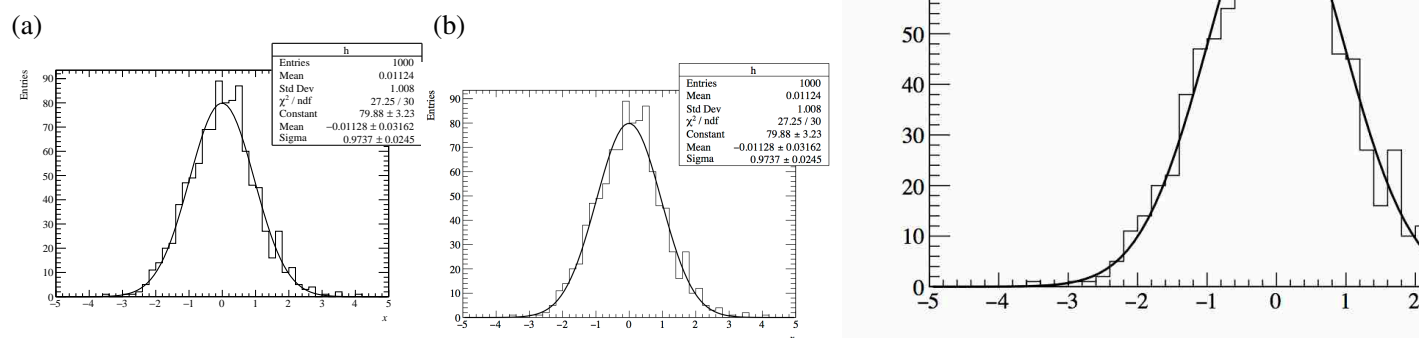


Fig. 4.2.4 異なる画像形式の比較。(a) PDF 形式。拡大しても綺麗であり、文字も検索やコピーができる。(b) PNG 形式。拡大するとビットマップ画像であることが分かる。文字を選択できない。(c) JPEG 形式。PNG に比べ、JPEG 圧縮特有のブロックノイズ、モスキートノイズが発生しており非常に汚いことが分かる。

Table 4.3.1 CTA で使用される望遠鏡の性能諸元

	大口径望遠鏡 Large-Sized Telescope (LST)	中口径望遠鏡 Medium-Sized Telescope (MST)	SC 中口径望遠鏡 Schwarzschild–Couders MST (SC-MST)	小口径望遠鏡 Smaller-Sized Telescope (SST)		
		FlashCam NectarCAM		GCT	ASTRI	1M-SST
エネルギー範囲	20–200 GeV	100 GeV – 10 TeV	200 GeV – 10 TeV	5–300 TeV		
台数 (北半球)	4	15	0	0		
台数 (南半球)	4	24	24	70–90		
鏡直径	23 m	12 m	9.7 m	4 m	4 m	4 m
焦点距離	28 m	16 m	5.6 m	2.3 m	2.15 m	5.6 m
視野	4.5°	7.7°	8°	8.6°	9.6°	9°
光学系	放物鏡	Davies–Cotton (DC)	Schwarzschild–Couders (SC)	SC	SC	DC
画素数	1,855	1,764 1,855	11,328	2,048	1,984	1,296

Table 4.3.2 表 ?? の悪い例

	大口径望遠鏡 Large-Sized Telescope (LST)	中口径望遠鏡 Medium-Sized Telescope (MST)	SC 中口径望遠鏡 Schwarzschild–Couders MST (SC-MST)	小口径望遠鏡 Smaller-Sized Telescope (SST)		
		FlashCam NectarCAM		GCT	ASTRI	1M-SST
エネルギー範囲	20–200 GeV	100 GeV – 10 TeV	200 GeV – 10 TeV	5–300 TeV		
台数 (北半球)	4	15	0	0		
台数 (南半球)	4	24	24	70–90		
鏡直径	23 m	12 m	9.7 m	4 m	4 m	4 m
焦点距離	28 m	16 m	5.6 m	2.3 m	2.15 m	5.6 m
視野	4.5°	7.7°	8°	8.6°	9.6°	9°
光学系	放物鏡	Davies–Cotton (DC)	Schwarzschild–Couders (SC)	SC	SC	DC
画素数	1,855	1,764 1,855	11,328	2,048	1,984	1,296

4.4 数式の使い方

L^AT_EX を使う理由のひとつが、数式を綺麗に出力できるというのがあります。例えば中性パイ中間子 π^0 のガンマ線への二体崩壊であれば

$$\pi^0 \rightarrow \gamma + \gamma \quad (4.4.1)$$

のように書けますし、もっとややこしい数式も色々と書けますが、詳細は「LaTeX 数式」などでインターネット上で検索してください。この例のように、本文中に数式を入れるときは $\$$ でその式を囲み、独立した行に数式を書くときは `equation` や `align`*2環境を使ってください。

4.4.1 斜体と立体

数式を書くときには「斜体」(italic) と「立体」(upright) の違いに気をつけてください。基本的に数式は斜体を使って書きます。何も考えずに L^AT_EX を使えば全て斜体になります。

ただし、次の 2 つの式を見比べてみてください。

$$e^{ix} = \cos x + i \sin x \quad (4.4.2)$$

$$e^{i x} = \cos x + i \sin x \quad (4.4.3)$$

式 ?? は全ての文字が斜体で書かれていますが、式 ?? は x 以外は立体です。このように、いくつかの文字では立体

*2 `amsmath` パッケージで使用可能です。

を使うのが一般的です。例えば \log 、 \sin 、 e （自然対数の底）、 i （虚数単位）、 d （微分作用素）などは、それぞれ `\log`、`\sin`、`\mathrm{e}`、`\mathrm{i}`、`\mathrm{d}` などと入力することで書くことができます*3。

ここで `\mathrm` というコマンドが出てきましたが、これは数式中で文字を立体にするためのコマンドです。特定の文字を立体にするときだけでなく、変数名の添字を立体するときにも使います。例えばトリガー回数を示す変数は N_{trigger} などと書くことがあると思いますが、このときに「trigger」の部分は変数ではありませんので、斜体にしません。

4.4.2 単位

数式中に単位を使うとき、`\mathrm` を使わずに 100MeV などとしてしまう間違いもよく見られます。このように斜体になったものは変数 M と e と V の掛け算であり、単位ではありません。また 100MeV のように単位と数値の間にスペースのない書き方をする人も見かけますが、これも間違いです。本文中に書くときは $100\sim\text{MeV}$ とし、`equation` 環境中では `100\ \mathrm{MeV}` と書きます*4。

\LaTeX では % の後ろをコメントとして扱いますので、95% のようにパーセントの表示をしたい場合には `95\%` のように書きます。% と数値の間にスペースを入れるかどうかは、流派が2つありますが、私の周りでは入れない人が多いようです*5。

4.5 引用の仕方

研究や論文というのは過去に誰かのやった研究を前提として新たに何かを進歩させるためにあります*6。しかしあなたの修士論文に全ての過去の研究を書くことはできませんので、引用という形式を使い他の論文をその事実の出典とします。

ここで、「引用」と日本語で書いた場合には「quotation」と「citation」の2つの英語に翻訳され得ますが、我々の論文で通常用いるのは「citation」のほうです。著作権法などで問題になるのは「quotation」のほうなので、間違えないようにしてください。

\LaTeX で `citep` コマンドや `citet` コマンドを使って論文を引用 (cite) するときは、例えば次のようになります。

宇宙線の全粒子スペクトルは図 XX に示すように 10^9 eV から 10^{20} eV までおよそ -3 乗の幂で減少している(?)。 10^{12} eV (1 TeV) 付近のガンマ線は超高エネルギーガンマ線と呼ばれ、様々な観測手法が提案されている(例えば ? を見よ)。この?の手法では...

ここでは引用 (cite) を3回しており、それぞれ `citep`、`citep`、`citet` コマンドを使っています。

4.6 BibTeX の使用

このテンプレートの場合、??ページに「引用文献」という箇所があります。このページを手作業で間違いなく整形するのは面倒です。手でやる代わりに \BibTeX という仕組みを使います。`thesis.bib` というファイルに引用文献の必要な情報が書かれていますので、これを参考にして \BibTeX ファイルを作るか、論文をダウンロードすると

*3 自然対数の底や虚数単位の場合は、分野や国によって斜体にするかどうかの違いがあります。また微分作用素は斜体で d とする場合もありますが、立体にすることで長さを表すのに頻繁に使われる変数 d と区別する効果があります。

*4 余計なバックスラッシュとスペースは、数字と単位の間スペースを入れるためです。

*5 入れない理由としては、% は単位ではなく 0.01 という数だから、というものが挙げられます。

*6 「巨人の肩の上に立つ」とよく表現されます。

きに `.bib` ファイルもダウンロードできますので、それを使ってください*7。

4.7 ヨーロッパ圏の人名など

ウムラウトなどの混じったヨーロッパ圏の人名を入力するには、例えばシュレーディンガーの場合、L^AT_EX では `Shr\{"o}dinger` と入力することで `Shrödinger` と表示すると L^AT_EX の教科書には書いてあります。しかしいちいちこんなことをするのは面倒ですので、`main.tex` に書いてある `\usepackage[utf8]{inputenc}` を使うことで、直接ウムラウト付きの文字を L^AT_EX のソース中に書いてしまって問題ありません。「ö」と「œ」は、この L^AT_EX ソース中では違う入力方法で書かれていますが、出力は同一です。

4.8 newcommand

入力が長く、論文中で何度も繰り返し使うような入力はコマンドとして登録することができます。例えば `\HI{}` や `\bs{}` といったコマンドを `main.tex` で定義しており、これらの結果は「HI」や「\」と表示されます。

*7 近年は「文献管理ソフト」と呼ばれるものが発達していますので、特に博士進学する学生は好きなものを入れてみてください。

第 5 章

剽窃について

5.1 剽窃とは何か

「剽窃（ひょうせつ）」とは

- ・「他人の作品や論文を盗んで、自分のものとして発表すること。」『大辞泉』
- ・「他人の作品・学説などを自分のものとして発表すること。」『スーパー大辞林』
- ・「他人の著作から、部分的に文章、語句、筋、思想などを盗み、自作の中に自分のものとして用いること。他人の作品をそっくりそのまま自分のものと偽る盗用とは異なる。」『ブリタニカ国際大百科事典 小項目事典』

のように辞書では説明されています。

例えばここで『ブリタニカ国際大百科事典 小項目事典』を引用元として明記せずに、

剽窃（ひょうせつ）とは、他人の著作から、部分的に文章、語句、筋、思想などを盗み、自作の中に自分のものとして用いること**です**。他人の作品をそっくりそのまま自分のものと偽る盗用とは異な**ります**。

という説明をしたとします。これが剽窃です。この例では赤字で示したとおり、文体をですます調に変更したり、読点を「,」から「、」に変更したり、文頭に「剽窃（ひょうせつ）とは、」と書き加えたりしていますが、全体としては同一の文章であるため、通常は剽窃と見なされます。

学術論文ではない創作物の形態によっては、剽窃行為が「インスパイア」や「オマージュ」という言葉で括られることもあります。しかし修士論文での剽窃行為は不正行為です。試験でのカンニングやレポートの丸写しと同じであり、（まともな大学や研究室であれば）厳しく罰せられます。

5.2 剽窃をするとどうなるか

修士論文中に剽窃行為が発見された場合、その学期における単位をすべて没収され、卒業に必要な単位が与えられず修士課程を修了できなくなる可能性が高いです。各大学や研究科でどのような対応を実際取るかはそれぞれだと思いますが、少なくとも私が審査員を担当した場合には落第させます。

修論審査に落第すれば、もし就職が決まっても留年を余儀なくされます。留年を選択せず修了を諦めて中退するにしても、就職先は剽窃行為のせいで修了できなかった学生をそのまま採用はしてくれないでしょう。仮に同じ企業に就職が認められたとしても、修士卒扱いで入社できたはずのところが学部卒扱いとなり、初任給が月額数万円低い状態から開始となります。例えば同期と 2 万円の月給差を保ったまま 40 年間働くとする生涯収入で 1000 万円程度の損失になります。もし留年する道を選んでも、定年時点で 1000 万円程度の年収を見込めるのであれば、生涯収入としてその額だけ失うことになります。

もし博士課程に進学する場合、なぜ留年したかの説明を陰に陽に常に求められます。たとえ直接にその理由を問われることがなくとも、他の学生より 1 年多く修士課程に時間がかかったということは、優秀な学生ではないと周りから見なされ、研究をする上でも奨学金などを取得する上でも不利になるでしょう。また標準年限を超えての在籍の場合、大学院の授業料免除などの制度も利用できなくなる可能性があります。

5.3 修士論文における剽窃について

節??に引用した一般的な剽窃の定義ではなく、科学文書や、特に修士論文での剽窃についてもう少し踏み込んで説明し直してみましょう。

5.3.1 いわゆるコピペ

少なくとも宇宙物理学分野における修士論文は独自性のあるものでなくてはなりません。独自性のある（オリジナル）とは次のようなことです。

- 誰かが過去にやった研究ではないこと
- 自分自身の手でやった研究であること（共同研究であれば、十分に自分の貢献のあること）
- 研究本体以外の章も含め、すべて自分の言葉で説明できること

したがって、誰かの論文や教科書の記述をそっくりそのまま持ってきて（いわゆる「コピペ」して）、それを自分の修士論文として提出することは許されません。高校や大学のレポートなどでも、他人のレポートを写すなど散々注意されるのと同じことです。

これはコピペする文章の長さに依りません。たとえ 1 行であってもコピペはコピペであり、剽窃と見なされます*¹。

もちろん、ある文章を他の論文や書籍から引用（quote）する必要がある場合は、逆に改変してはいけません。そっくりそのまま書き写し、それを自分の文章とは別のものであると分かるように引用符や枠で囲むなりします。しかし宇宙物理学関連の修士論文でこのような引用をすることは、ほとんどないと思います。

5.3.2 他人の文章の改変

コピペとともによく見られるのが、他人の文章を一部だけ改変して自分が書いたかのように装うことです。完全に同一のものを持ってくる方が簡単ですし、なぜこのような行動を取るのかよく分かりませんが、私の経験として最も多い剽窃行為がこの文章の一部改変です。

もしかすると「先輩の修論を写したりコピペするなよ。自分の言葉で書けよ」とだけ教員から指導を受けると、表面的に一部改変すれば剽窃にはならないと勘違いするのかもしれませんが。しかし元の文章が存在しなければ作成できないのですから、これは独自性のある文章とは見なされず、やはり剽窃行為となります。

たとえば次のような文章が「元ネタ」として存在していたとしましょう*²。

1910 年代に Hess らによって宇宙線の存在が確認されて以来、様々なエネルギー領域、様々な検出器によって宇宙線の観測が行われてきた。同時に、ガリレオ以来発達してきた可視光による天体の観測も、電波望遠鏡や赤外望遠鏡の登場によって多波長での観測へと発展することとなった。

宇宙線と言っても、その成分は電磁波、陽子、原子核、neutrino など様々であり、それらの持つエネルギー

*¹ ただし、ごくありふれた表現や、酷似するのが避けられない科学的事実を除く。

*² これはきちんと添削を受けていない、今となっては恥ずかしい私の修論の一節ですが、あくまで例です。

も広範にわたる。現在地球上で確認されている宇宙線のうち、最もエネルギーの高いものは 10^{20} eV を超える（最高エネルギー宇宙線）。これは人工的に到達できるエネルギーを実に 8 桁も上回るが、なぜそのような高エネルギーの宇宙線が存在するのかは謎である。加速機構、地球までの伝播過程、1 次宇宙線成分は何であるのか、いずれも未解明のままであり、その興味は尽きない。

?より引用

少しこれを改変してみましょう。赤字が削除箇所、青字が追加箇所です。実際に私が発見してきた剽窃行為には、このような改変が多くありました。

1910年1912代年に Hessらによって宇宙線の存在が確認初めて発見されて以来、様々な広いエネルギー領域範囲、様々な多種多様な検出器によって宇宙線の観測が行われてきた。同時にまた、ガリレオ以来発達してきた可視光による天体の観測での天体観測も、電波望遠鏡や赤外望遠鏡という新しい観測手段の登場によって多波長での観測により、多波長観測へと発展することとなったした。

宇宙線と言いつても、その成分は電磁波、陽子、原子核、neutrino電子、ニュートリノなど様々であり、それらの持つそのエネルギー範囲も広範何桁にもわたる。現在地球上で確認されている宇宙線のうち、最もエネルギーの高いものは 10^{20} eV を超える（いわゆる最高エネルギー宇宙線）。これは人工的に加速器で人類が到達できるエネルギーを実に 8 桁も上回るが、なぜそのような高いエネルギーの宇宙線が存在するのかは謎である解明されていない。宇宙線の加速機構、地球までの伝播過程、また 1 次宇宙線成分は何であるのかは、いずれも未解明決のまま問題であり、その興味は尽きない将来の宇宙線観測計画による解決が期待される。

?を意図的に改変

5.3.3 元の文章を下敷きに自分で考えたつもりになったもの

さらに改変の量を増やし、ところどころに自分の独自の文を入れたり、文の前後を入れ替える剽窃もあります。自分で考えて文を挿入するのだから剽窃ではないと考える人もいるかもしれませんが、やはり元の文章が存在しなければ書くことのできない文章ですので、これも立派な剽窃です。たとえば次のようなものです。

Hess の気球実験によって 1912 年に宇宙線が大气中で発見されてから、様々な粒子、多様な検出手法、また MeV 領域から 10^{20} eV にまでおよぶエネルギー範囲で宇宙線の観測が行われてきた。一方、電磁波による天体の観測も、ガリレオによる可視光観測に始まり、電波望遠鏡や赤外線望遠鏡などの登場によって他波長観測へと発展した。さらに近年の重力波やニュートリノ観測を加え、現在の宇宙観測は、多粒子、他波長観測の時代、すなわちマルチメッセンジャー天文学へと進展した。

このうち宇宙線は、陽子、原子核、電子、ニュートリノなどを含む、宇宙空間を飛び交う高エネルギーの粒子である。先に述べたように、その最高エネルギーは 10^{20} eV にまでわたる（いわゆる最高エネルギー宇宙線）。これは人類が LHC 加速器で到達できる数 TeV というエネルギーを 8 桁も上回るものであるが、なぜそのような高いエネルギーの宇宙線が宇宙で加速されているのか、宇宙線の発見から 100 年以上が経っても未解決の問題である。その加速機構、加速天体、地球までの伝播、また粒子の種類がなんであるかという謎を解き明かすには、今後の宇宙線観測手法に大きな飛躍が必要である。

?を意図的に改変

ここまで改変すると、全く違う文章のように感じる人もいるかもしれませんが、実際に行われる剽窃行為では、

このような元ネタに改変を加えた文章が何段落も続くことが多いです。そのため、文章の一部が似通っているだけでなく、その章の論理展開自体がほとんど同じになってしまうのです。

研究背景は過去に行われた研究の積み重ねなので、論理展開が同じになることは仕方がないという主張をする学生もいます。しかし修士論文はその研究目的が各々違うわけですから、論文のイントロなどで全く同じ論理展開になることは本来ありえません。その論文独自の研究内容を説明するためにイントロは書かれるべきであり、他の文章と同じであるというのは、イントロを書くという目的を勘違いしています。

5.3.4 出典のない図表の使用

他人の文章を剽窃する行為とは別に、図表を適切に引用（cite）せずに流用するという剽窃もあります。これは悪意があって行われているわけではなく、引用の作法を知らないだけの場合が多いため罪としては軽いかもしれませんが、しかし、その修士論文の読者に対して「この図は自分が作りました」と嘘をつくのと同じ行為ですので、やはり問題行為であることは理解できると思います。

このような図表の剽窃は、特に共同研究で多く見られます。ある実験プロジェクトに参加している場合、実験装置の説明の図や写真をプロジェクト内で使いまわすことがあるでしょう。たとえば図??のようなものが該当します。もしこれを出典もしくは作者を明記せずに使用した場合、剽窃行為に当たります*3。

図表の提供者の名前を入れる、その図が最初に使われた論文や出版物が存在する場合はそれを出典として明記する（cite する）、もしくは提供した実験グループなどの名前を入れるなどしてください。

5.3.5 アイデアの盗用

他人の考えた研究アイデアを自分が考えたかのように記述するのも剽窃です。例えば投稿論文になっていないものの、先輩の修士論文で先行研究が行われていたとしましょう。これを先行研究として取り上げることなく、「〜という手法を本論文では考案し」などと書くのは剽窃行為です。きちんと「〜という手法が先行研究で提案され、本論文ではこれを発展させ」のように書きましょう。

5.3.6 自己剽窃

自己剽窃とは、自分の書いた論文などから図や文章を剽窃して再利用することです。なぜこれが問題とされるのか、直感的にはすぐに分からないかもしれません。

自己剽窃が最も問題とされるは、論文の二重投稿です。どこかで論文を出版する場合、レビュー論文でない限り、それぞれが独自の新規性を持つ論文でなくてはなりません。したがって、業績稼ぎのために同じ内容の論文を複数の場所で発表するのは研究不正として扱われます。

次に自己剽窃が問題となるのは、著作権の問題です。投稿論文を科学誌に掲載する多くの場合、その著作権を出版社に譲渡することになります。最近のオープンアクセス（open access）誌の場合には著作権が論文著者に残される場合もありますが、投稿論文の著作権を必ずしも自分が持っているわけではないのだということを覚えておいてください。

著作権が出版社にあるということは、その著作物を引用の範囲を超えて勝手に再利用してはいけないということになります。著作権、英語で書くと copyright ですが、すなわち複製する権利を出版社に譲渡してしまっているからです。

ただし、多くの出版社では学位論文や国際会議のプロシーディングスなどで、著者が図表などを出版社

*3 おそらく「出典を明記して再提出しろ」と言われるだけで、落第はしないと思いますが。

に断らずに使いまわすことを許可しています。ただし、出典を明記することは求められていることが多いはずです。もし投稿論文に使用した図表もしくは文章を修士論文で使いまわす場合、出版社との著作権の契約について理解しておきましょう。たとえば Elsevier 社の場合、<http://jp.elsevier.com/authors/author-rights-and-responsibilities> に著者の権利が書かれています。他の出版社も同様の情報を公開しています。

5.4 なぜ剽窃は許されないのか

なぜ剽窃行為は許されず、それが修士論文で不正行為とされるのか、その理由を改めてまとめます。

1. 学位審査は、学生が研究背景などを理解しているか、またそれを自分の言葉で伝える能力を身につけているかを審査する場です。したがって、剽窃を含む文書ではこの審査を適切に行えなくなってしまいます。修士の学位を与える審査の一環として修士論文を執筆しているわけですから、修士論文作成能力がないのにそれを他人の文章を使って誤魔化すのは、当然不正行為になります。
2. 同じ文章を使いまわすとき、一般的には引用 (cite ではなくて quote) をし、自分の書いた文章と他人の文章を区別するのが標準的です。超新星の過去の記録など一部の例を除き、宇宙物理学分野で quote のほうの引用をすることはほとんどありません。もし必要となる場合は、他人の書いた文章であることが明確に読者に分かるようにしましょう。自分で作った文章かのように見せるのは決して許される行為ではありません。
3. 他人の書いた文章を自分が書いたかのように見せるのは、人の手柄を横取りすることになります。
4. 少なくとも日本の国内においては、他人の著作物を勝手に使用したり改変したりすることは、著作権の侵害に当たる行為です。
5. 元の文章を無理に改変することにより、推敲された元の文章よりも質の低い文章になることが多く、また間違った記載となる場合が多々あります。例えば「突発天体を観測する」を無理やり「突発天体を監視する」に変更することにより、意味が大きく変わることもあります。
6. 同じものを繰り返すというのは、先人の研究をさらに発展させていくという、科学の営み自体を否定する行為です。
7. 過去数年で該当分野に大きな進展があった場合にも、それを無視した様な文章が生産されてしまいます。例えば 2018 年の修論なのに重力波が未だ検出されていない前提の文章になっていたりということが考えられます。
8. 修論の添削をする教員は、執筆した学生の研究能力や文章作成能力を高めるために添削をしています。良い出来の修論を書かせることが目的ではないのです。そのため、本人が書いてすらいらない文章を添削させ、大学教員の貴重な時間を奪うことは、学生と教員の間の信頼関係を大きく毀損する大変失礼な行為です。またそのような添削をしても本人が書いていないのですから、その学生の能力向上には全く役に立たず、学生も自分で考えることなく言われるがままに改訂を繰り返すことになるでしょう。

第 6 章

議論

ここではこの研究で得られた結果についての議論を行います。測定結果や観測結果などと一緒に議論を進める場合もあるので、必ず必要な章であるとは言えませんが、できる限り研究で得られた事実と自分の議論は分けましょう。

第 7 章

結論

ここには自分の修士論文の結論を書きます。「議論」の章で書かれたことも、再びここに短く書かれます。

「序論」で始めたら「結論」、「はじめに」で始めたら「おわりに」が原則です。ただし、「まとめと今後の展望」などとすることもありますので、好みに応じて変えてください。

付録

「付録」(appendix)は、論文の本文に載せるには情報として邪魔もしくは必須ではないものの、読者にとって有益となるような情報を載せます。付録を必要としない論文もちろん存在しますので、そこは著者の判断です。

例えば、たくさんの観測データを様々なモデルでフィットした場合、フィット結果の絵がたくさん出てくるはずで、そのような図は本文中に大量に出されても大切な情報を見失ってしまいますので、大部分は付録に載せることが推奨されます。他には、何かしらの長い式変形や証明を載せる必要がある場合、付録に移動する場合があります。

A すごい長い証明

式(??)のように、式番号がアルファベットとアラビア数字の組み合わせになるように、`LATEX`ソース中で設定してありますので、中身を眺めてみてください。

$$1 + 1 = 2 \tag{A.1}$$

B すごいたくさんのフィットの図

謝辞

「謝辞」(acknowledgments)は、修士論文を作成する上であなたを支えてくれた人への感謝を書く場所です。誰かへの感謝の気持ちを公開の場所で文書にするというのは気恥ずかしいものですし、修士論文以外ではそんなことをした経験がないかもしれませんが、投稿論文では一般的に行われます*¹。

多くの修士論文では指導教員(国立大学の法人化後は、指導教「官」とは言いません)、実験協力してくれた共同研究者、間接的に助言などをくれた研究室の他の教員・先輩・同輩・後輩、研究室の秘書さんなどに謝意を示すことが多いようです。もし奨学金をどこからか受給していたら、奨学金の出所に対しても謝辞を書いても良いでしょう。また家族・恋人・友人に対する感謝も見られますが、恋人の名前は将来隠したくなる場合もあるので注意しましょう。

感謝の気持ちを書く場所ですので、その相手に失礼のないようにしましょう。氏名の漢字の間違いや、職階の間違いが頻繁に見られます。特に助教を助教授と書き間違えたり、准教授を助教授としたりという間違いが目立ちます。

*¹ ただし、投稿論文では家族や友人への感謝はあまり書かず、研究費を出した機関や研究の協力をしてくれた研究者などを書くことが多いです。