

AIするディープラーニング

AI Love Deep Learning

b1015031 黒川大智 Daichi Kurokawa

1 背景と目的

2010年代に突入してから今日に至るまで、多層ニューラルネットワークを用いた機械学習手法「ディープラーニング（深層学習）」の研究は急速に発展してきた。現在では画像処理・音声処理・自然言語処理などの幅広い分野に応用が進み、いずれにおいても高い性能を示している。また、企業や大学などの研究組織がライブラリを公開したことにより、我々の身近な分野にもディープラーニングが活用されるようになってきている。例えば、FacebookはDeepFaceという顔識別システムによって97.25%の精度でヒトの顔を見分けることに成功^[1]しており、この技術はFacebookに投稿された人物写真の顔識別などに応用されている。また、Appleの音声アシスタントシステムSiriは、音声認識を2014年からディープラーニングによって行っている^[2]。さらに、Microsoftのチャットボット「りんな」は自然言語の学習にディープラーニングを活用している^[3]。

以上のように、ディープラーニングの活用範囲は広範に及び、今もなお拡大し続けている。しかし、新しい分野であるがゆえに未だ注目されていない領域も多い。そこで本プロジェクトでは、先行研究にない新しいディープラーニングの応用領域を開拓することと、その領域での課題解決を行うこと目的として活動をおこなった。

2 基礎知識の習得とテーマ設定

4月～5月は、主に基礎的な知識習得と先行研究調査に取り組んだ。

基礎的な知識として、Pythonによるプログラミングとディープラーニングの初歩を学んだ。Pythonは機械学習に関するライブラリが充実しているため、本プロジェクトで主に使用する言語として採用した。プログラミングの学習には、オンライン学習環境である「Progate」を採用した^[4]。ディープラーニングの初歩

についての学習は書籍^[5]を用いて行った。

先行研究調査は論文や書籍、Webなどの情報を参考にプロジェクトメンバ各個人で行った。その後、全員でブレインストーミングを行い、グループテーマとして設定可能と思われる領域を選考した。その結果、グループAでは「手話」、グループBでは「詰将棋」を対象に課題解決を行うこととなった。

3 課題解決のプロセスとその結果

本節では各グループにおける課題解決のプロセスと、実験の結果、考察について報告する。

3.1 グループA：CNNを用いた指文字の認識

3.1.1 先行研究

手話を母国語へ翻訳するシステムにはいくつかの先行事例がある。ひとつはNavidらが開発した「SignAloud」である^[6]。このシステムは、手の位置と動きを検出する特殊なグローブを使用することにより、手話話者が示したジェスチャを音声として出力することができるものである。また舩阪ら^[7]の研究では、ジェスチャ入力装置であるLeap Motionを用いて87種類の指文字を認識させることに成功している。

しかし、これらの研究ではジェスチャを認識するために何らかの特殊装置を用いている。このため、手話話者はシステムを用いる際に装置を別途用意する必要があり、また、その装置の有効範囲のみでしか口語話者とのコミュニケーションがとれないという課題点がある。

3.1.2 映像から直接翻訳することのメリット

本グループが提案する映像からの直接翻訳には、以下に示すような2点のメリットが存在する。1点目のメリットは、スマートフォンやノートPCに搭載されているカメラで手話を読み取れる点である。先行研究では独自に開発したグローブや特殊なセンサを用いているが、カメラは多くのスマートフォンやノートPCに標準搭載されているため、それらさえあれば手話の読み取りを実行することができる。2点目のメリットは、手話認識の

有効範囲をユーザが調節可能であるという点である。先行研究では、「グローブが接続されている Bluetooth の有効範囲内」や、「Leap Motion の有効範囲内」でのみしか手話の認識ができない。装置の有効範囲は目に見える形では分からないため、手話話者は有効範囲からの逸脱に気づきにくいということが考えられる。カメラの場合は、手話話者がカメラの枠内から逸脱しないように調節しながら手話を認識させることが可能である。そのため、手話翻訳が途中で途切れるなどのトラブルを未然に防ぐことができるというメリットがある。

3.1.3 手話の学習における問題と解決する課題

ディープラーニングによって映像から手話を翻訳するためには、解決しなくてはならない問題が主に 2 点ある。1 点目の問題は、顔や手など複数の情報を関連付けて翻訳しなければならない点である。手話の中には腕や手の動きと顔の表情を組み合わせることによって意味を成すものがある。よって映像から手話を翻訳する際は、表情と手の動きを同時に読み取って翻訳しなくてはならない。2 点目の問題は、ディープラーニングを行う際に手話翻訳に必要な情報のみを抽出してする必要があるという点である。手話において意味を表しているのは手と顔の情報であり、それ以外の情報は無駄な情報といえる。ディープラーニングによって学習を行う際、無駄な情報は学習に悪影響を及ぼす可能性があると考えられた。従って、ディープラーニングを行う際には無駄な情報を極力排除した映像を用いる必要がある。

以上 2 点の問題のうち、1 点目の問題はシステムの実用上の問題であり、2 点目の問題は手話の学習（特徴抽出）に関する問題であると考えられた。本年度はより基礎的な、2 点目の特徴抽出における問題の解決を行うこととなった。題材として、手話の一種である指文字を取り上げ、映像中から翻訳を行うことができるのかを実験により確かめることとした。

3.1.4 提案するアイデア

指文字の学習における特徴抽出の問題を解決するためには、本グループでは映像中の肌色要素の抽出を行った。肌色要素の抽出を行うことで、ディープラーニングを行う際に不必要な被写体の服や背景の情報を排除することができた。この肌色抽出処理を学習に用いるすべての画像に対して行った。

ディープラーニングの具体的な手法としては、画像

認識タスクに専ら用いられる畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) を採用した。CNN は画像認識タスクの中でも、1 枚の画像からそこに写る物体のカテゴリ名を識別する物体カテゴリ認識に有効である^[8]。本グループが行う指文字の認識も物体カテゴリ認識に該当すると考えられるため、CNN の採用は妥当であると考えられる。

3.1.5 実験

本グループでは入力層 (1)、中間層 (4)、出力層 (1) の計 6 層の CNN を構築し、指文字の学習および認識と翻訳に関する実験を行った。

CNN の学習を行うために用意したデータセットは、「あ」行から「た」行までの指文字画像 84000 枚と「その他」画像 15000 枚である。本来なら 50 音すべての指文字データセットを用いて実験するべきであったが、学習を行うための CNN 構築が予想以上に難航したため、「あ」行から「た」行までの認識実験を行うにとどまった。「その他」画像は、「あ」行から「た」行以外の画像を分類できるかを調べるためのダミーデータである。

上記に示したデータセットのうち 80400 枚を訓練データ、残りをテストデータとして用いた。CNN の訓練は 50 エポック行った。訓練の結果得られた学習済みモデルを用いて、テストデータに対する汎化性能を調べた。結果として CNN の訓練データに対する正解率とテストデータに対する正解率を出力させた。

3.1.6 結果と考察

学習済みモデルの正解率は、訓練データにおいて 98% 以上、テストデータにおいて 71~78% であった。訓練データにおける正解率とテストデータにおける正解率との間には 20% 程度の差があった。このことから、学習済みモデルはテストデータに対する汎化性能を十分に得られていないと考えられた。その理由としては、データセット画像でサンプルとして用いた人物が 12 人程度と少なかったことや、データセット数を増加させるための「ぼかし」や「ずらし」などの加工が十分ではなく、同じような画像が訓練に用いられたことなどが考えられる。

3.2 グループ B : Word2vec を用いた詰将棋問題の生成

3.2.1 先行研究

詰将棋問題の創作をコンピュータで実現しようとする試みは 1990 年代からある。それらの試みの中で本グ

グループが目したアルゴリズムは、野下^[9]によるランダム配置を用いた手法（ランダム法）である。このアルゴリズムは、ランダムに駒を並べた配置から始めて、そこから駒の削除、再配置を繰り返しながら詰将棋を創作していくものである。このアルゴリズムのメリットは、主に13手から長くて19手詰め以上の詰将棋を生成できる点である。一方で、単に駒を打ち込むだけで詰みが得られたり、一見しただけで詰みが分かっていたりなど、簡素で面白みに欠ける詰将棋が生成されやすいというデメリットも存在する^[10]。本グループでは以上に示したランダム法をベースに、面白みのある詰将棋の生成に取り組むこととした。

3.2.2 検討した機械学習手法

詰将棋をディープラーニングで学習させるにあたり、いくつかの手法を検討した。検討した手法は主に「画像処理的なアプローチ」と「自然言語処理的なアプローチ」とに大別された。画像処理的なアプローチとは、盤面の1マス1マスを画像のピクセルのように捉えることで、盤面全体を9×9の画像として扱うというアプローチである。一方自然言語処理的なアプローチとは、詰将棋の解手順（詰め手順）を文章のように捉えることで、自然言語と同じように詰将棋を扱うというアプローチである。

検討の結果、自然言語の学習と詰将棋の学習とは類似する点が多いと考えられたことから、本年度は自然言語処理的なアプローチで詰将棋の学習を行うこととなった。具体的には、文章における単語の意味推測に用いられる機械学習手法 Word2vec を用いて、詰将棋における駒の意味や前後手との関係性を学習させることとなった。

3.2.3 提案するアイデア

本グループが提案する詰将棋生成のアイデアは以下である。

(1) Word2vec を用いて、人間が作った詰将棋の詰め手順から各手の意味合いを学習させた学習モデル A を作成する

(2) ランダム法を用いて詰将棋を生成し、その詰め手順を抽出する

(3) (2) で抽出した詰め手順の1手1手を学習モデル A に入力する。学習モデル A は、入力された手に意味合いが近い候補手を出力する

(4) 出力された候補手が入力した手と入れ替え可能であれば、その手を入れ替える

以上の手順で生成された詰将棋は、Word2vec の学習モデルのアドバイスを参考に作った詰将棋となる。Word2vec の学習モデルは、人間の詰将棋から駒や指し手の意味合い・狙いをモデルとして獲得していると考えられる。従って、モデルの出力した候補手と入れ替えが成功すれば人間が作った詰将棋に近い詰将棋を生成できると考えられる。このような、「詰め手順」から詰将棋の「面白さ要素」が獲得できるという仮説の元、本グループでは以下に示す実験を行った。

3.2.4 実験

実験に用いたデータセットとして、これまで新聞や雑誌などに掲載された詰将棋 14000 局を用いた。元となった詰将棋は全日本詰将棋連盟の荒川様にご提供いただいた。

データセットから抽出した「詰め手順」をスペース区切りでテキスト化し、これを元に Word2vec の訓練を行った。実験に用いた Word2vec は、自然言語処理において利用されていたネットワークをそのまま用いた。自然言語における「単語」と詰将棋における「1手」は対応しており、特にネットワークに加工を行わなくても学習モデルの作成が可能と考えたためである。100 エポックの訓練の結果学習モデルを得た。

この学習モデルを用いて、詰将棋問題の生成を行った。はじめにランダム法を用いて基礎となる詰将棋を生成した。次にこの詰将棋の詰め手順を学習モデルに入力し、出力された候補手から入れ替え可能な手を選び入れ替えた。最後に、入れ替えた後の詰将棋を図化した。

3.2.5 詰将棋の評価

詰将棋の評価は、広瀬ら^[11]において用いられた評価関数を参考に行った。評価関数を用いることにより、詰将棋が面白くなったかどうかを定量的に評価することができる。評価関数は表 2 に示すような複数の指標を元に、詰将棋の「面白さ」を点数化するものである。この評価関数では、値が 0 に近づくほど詰将棋に対する評価が高い。

3.2.6 結果と考察

入れ替えを行う前と後で、詰将棋の「面白さ」に変化があったのかを評価関数を用いて評価した。その結果、入れ替えを行う前の評価値は-310 であったのに対して、

入れ替えを行った後の評価値は-232 となり、入れ替えを行う前より 78 ポイント評価値が高かった。それぞれの詰め手順を見ると、入れ替え前は持ち駒を打ち込むような単調な詰将棋であったが、入れ替えを行った後は打ち込む手の他に大駒を使った手も現れるようになった。詰め手順の数も入れ替え前の 7 手詰めから 9 手詰めに増えていた。

以上の結果を踏まえて、入れ替え後の詰将棋の評価が上がった理由を考察する。まず、評価関数が「面白さ」を評価する指標に「駒の重み」「手数に対する持ち駒の数」という項目があった。この項目は上述した「大駒の使用」や「手数の上昇」に関わるパラメータである。大駒は利き筋が広いため、使用頻度が上がれば詰将棋の複雑性も上がる。手数についても、増えれば増えるほど詰将棋を解くために思考力を要するようになる。従って、入れ替え後の詰将棋は入れ替える前に比べて思考力を要する複雑な詰将棋になったため、評価値が上がったと考えられる。以上の事実から、Word2vec を用いた手順入れ替えによる詰将棋の生成は、「面白い」詰将棋の生成に有効であると考えられる。

4 今後の課題

今後の課題として、グループ A では 50 音すべてに対応した指文字の認識・翻訳を行うことがある。今年度は技術的なノウハウがなかったため CNN の構築が難航し、「あ」行から「た」行までの認識にとどまってしまった。今回開発した指文字認識システムを実用的なものにするには 50 音対応であることが必須である。また、動きのある手話や顔の表情と関連した手話を翻訳できるようなシステムに改良する必要もあると考えられる。

グループ B では、人力による入れ替え作業をコンピュータによって自動化するという課題が残った。現状のシステムでは、学習モデルが出力した候補手との入れ替えはすべて人力に行っている。これが自動化できれば、候補手と入れ替え可能かを余すことなく探索できるため、さらに質のよい詰将棋を生成できる可能性がある。また、今年度は駒入れ替えが「面白い」詰将棋生成に有効かを確かめる実験を行ったのみで、ランダム法自体の改良は行っていない。ランダム法自体を機械学習により改良できれば、より効率よく質の高い詰将棋を生成することができると考えられる。

参考文献

- [1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, Deepface: Closing the gap to human-level performance in face verification, In *Proc. CVPR*, pp. 1701-1708, 2014.
- [2] Steven Levy, THE IBRAIN IS HERE—AND IT'S ALREADY INSIDE YOUR PHONE(2016), WIRED, <https://www.wired.com/2016/08/an-exclusive-look-at-how-ai-and-machine-learning-work-at-apple/>, 2016.
- [3] 太田智美, ついに明かされる「りんな」の“脳内” マイクロソフト、「女子高生 AI」の自然言語処理アルゴリズムを公開, ITmedia News, <http://www.itmedia.co.jp/news/articles/1605/27/news110.html>, 2016.
- [4] Progate, <https://prog-8.com/>
- [5] 斎藤康毅, ゼロから作る Deep Learning -Python で学ぶディープラーニングの理論と実装. オライリー・ジャパン, 2016.
- [6] University of Washington UW undergraduate team wins \$10,000 Lemelson-MIT Student Prize for gloves that translate sign language. <http://www.washington.edu/news/2016/04/12/uw-undergraduate-team-wins-10000-lemelson-mit-student-prize-for-gloves-that-translate-sign-language/>, 2016
- [7] 船坂真生子, 石川由羽, 高田雅美, 城和貴. Leap Motion Controller を用いた指文字認識. 情報処理学会研究報告, Vol.2015-MPS-102, No.8, 2015.
- [8] 岡谷貴之, 深層学習. 講談社, 2015.
- [9] K. Noshita, A Note on Algorithmic Generation of Tsume-Shogi Problems, In *Game Programming Workshop in Japan '96*, pp. 27-33, 1996.
- [10] 石飛太一. 詰め将棋問題の自動生成アルゴリズムに関する研究. 北陸先端科学技術大学院大学, 研究課題報告書, 2013.
- [11] 広瀬正幸, 伊藤琢巳, 松原仁. 逆算法による詰め将棋の自動創作. 人工知能学会誌, Vol.13, No.3, 1997.