

卒業研究報告書

題 目

遺伝的アルゴリズムによる機械学習における
疑似ラベル生成手法の提案

研究グループ 知能情報第1グループ

指導教員 森 直樹 教授

令和 2 年 (2020 年) 度 卒 業

(No. 1171201161) 細川 岳大

大阪府立大学工学域電気電子系学類情報工学

遺伝的アルゴリズムによる機械学習における 疑似ラベル生成手法の提案

第 1 グループ 細川 岳大

1. はじめに

近年、ラベル付きデータ生成に伴うコストの観点から分類問題において半教師あり学習 (Semi-Supervised Learning: SSL) と呼ばれる学習データ全体のうち一部にのみラベル付けされた状態で学習を行う手法が盛んに研究されている。しかし、当然のことながら各ラベルに対するデータ数が少ないほど精度が安定しなくなるという問題も報告されている。一方で、ラベルなしデータへのラベル付けタスクは一種の組合せ最適化問題と考えることができる。

そこで本研究では、組合せ最適化遺伝的アルゴリズムを用いてデータにラベルを付与することでラベル付きデータが少ない場合における半教師あり学習の頑健性を高める手法を提案する。

2. 要素技術

2.1. Contrastive Learning

Contrastive Learning (CL) とは、特徴表現を獲得するための自己教師あり学習のひとつである。一つの画像から得られる特徴表現が画像変換によって画像の持つ意味が大きく変わらない変形を獲得することができる。

2.2. SimCLR

Simple framework for Contrastive Learning of visual Representation (SimCLR) [1] は SSL の一つであり、CL により特徴抽出器を学習したうえで、少量のラベル付きデータによる Fine tuning によって分類器を獲得する手法である。

2.3. 遺伝的アルゴリズム

遺伝的アルゴリズム (Genetic Algorithm: GA) は生物の進化を模倣した最適解探索アルゴリズムである。選択、交叉、突然変異の 3 つの操作によって最適解を探索することができる。

3. 提案手法

本研究では、CIFAR-10 においてラベルありとラベルなしに分割し、ラベルなしデータの一部に対するラベルを GA によって探索する。GA の個体について各遺伝子はラベルデータのいずれかの整数値とし、取り出されたラベルなしデータに対する疑似ラベルとする。以下に提案手法の概要を示す。

1. CL によって特徴抽出器を学習する。
2. GA の個体と対応するデータを学習データとし、Fine tuning を行い、ラベル付きデータに対する識別率を適応度として GA の探索を行う。
3. 探索された個体とラベル付きデータを学習データとし、Fine tuning を行い、テストデータを識別する。

4. 数値実験

4.1. 実験方法

データセットとして 10 クラス識別問題である CIFAR-10 を用いた。学習に 50000 件、テストに 10000 件をそれぞれ使用し、ラベル付きデータ及びランダムに選ばれた疑似ラベルを付与する対象であるラベルなしデータはどちらも学習データから選んだ。CL の学習について、特徴抽出器には Resnet-18 を用い、学習データ 50000 件を、500 epoch 学習した。また分類器は MLP を使用した。GA の適応度計算については、30 epoch 学習した。

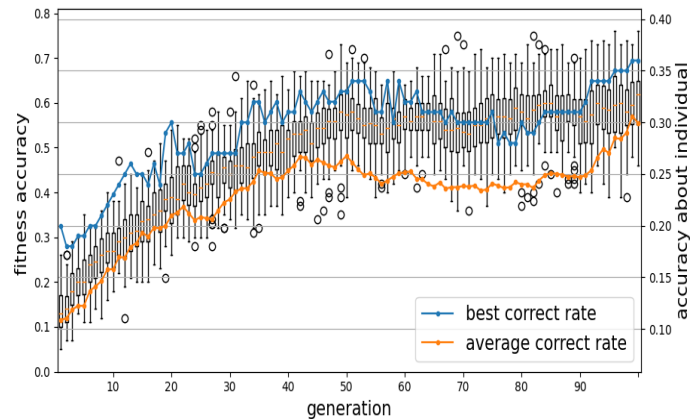


図 1: 各世代に対する適応度と疑似ラベルの正答率

4.2. 結果と考察

表 1 と図 1 に結果の一例を示す。また、図 1 の横軸に世代を、縦軸は箱ひげ図が適応度、折れ線グラフが各個体の疑似ラベルに対する正解ラベルの割合である。提案手法が最も低い精度となった一方で、GA の探索については進んでいるといえ、疑似ラベルに対する真の正答率は 37% まで上がっていることが確認できた。

考察として GA の探索において今回の設定では多様性を保つことができず、局所解に陥ってしまったのではないかと考えられる。また、ラベルの多くが誤ラベルのため train_accuracy が学習において 80% 以上にならず各個体に対しての正確な適応度が求められなかった可能性がある。

表 1: テスト識別率結果

生成ラベル	テスト識別率
baseline モデルによる疑似ラベル	0.784
提案手法による疑似ラベル	0.674
正解ラベル	0.822
baseline (ラベル付きデータのみ)	0.772

5. まとめと今後の課題

本研究では、GA によるラベルなしデータに対する疑似ラベルの生成手法を提案した。

今後の課題として、GA の拡張手法や適応度の計算方法を工夫することでより良質な疑似ラベルを生成し識別率向上を目指すことが挙げられる。

参考文献

- [1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

目次

1	はじめに	1
2	要素技術	2
2.1	半教師あり学習	2
2.2	疑似ラベル	2
2.3	FixMatch	2
2.3.1	RandAugment	2
2.3.2	Consistency Regularization	3
2.3.3	更新	3
2.4	SimCLR	4
2.4.1	Cotrastive Learning	5
2.5	ResNet	5
2.6	Genetic Algorithm	5
2.6.1	選択	6
2.6.2	交叉	7
2.6.3	突然変異	7
3	提案手法	8
3.1	個体設定	8
4	提案手法	8
5	数値実験	9
5.1	実験1	9
5.1.1	実験設定	9
5.1.2	結果と考察	12
5.2	実験2	14
5.2.1	結果と考察	17
6	まとめと今後の課題	19

目 次	ii
謝辞	20
参考文献	21

図 目 次

2.1	FixMatch の概略図	3
2.2	SimCLR の概略図	4
2.3	Residual Block の構造	6
5.1	実験 1 : GA の探索結果	12
5.2	実験 1 : 探索個体の散布図	13
5.3	実験 1 : GA の探索結果	17
5.4	実験 2 : 探索個体の散布図	18

表 目 次

5.1	実験 1 : データ内訳	9
5.2	実験 1 : GA の設定	10
5.3	実験 1 : FixMatch の設定	11
5.4	実験 1 : テスト識別率	14
5.5	実験 2 : データ内訳	14
5.6	実験 2 : GA の設定	15
5.7	実験 2 : SimCLR の設定	16
5.8	実験 2 : テスト識別率	18

1 はじめに

近年、機械学習の発展に伴い、様々な分野への応用がされており様々な新規データセットにおいて目覚ましい結果が報告されている。また新規データセットを生成する際、分類問題では各データにふさわしいラベル付けをする必要があるが、ラベル付けには人の手が必要でありコストがかかる問題がある。そこで半教師あり学習 (Semi-Supervised Learning: SSL) と呼ばれる学習データ全体のうち一部にのみラベルが付与された状態で学習を行う手法が提案されており、盛んに研究されている。また最新の SSL の手法では、全データにラベルが付与されている教師あり学習にも劣らない成果の報告^[1]もある。しかし、当然のことながら各ラベルに対するデータ数が少ないほど精度が安定しなくなるという報告もされている。

一方、ラベルなしデータへのラベル付けタスクは一種の組合せ最適化問題と考えることができる。

本研究では、組合せ最適化問題に有効である遺伝的アルゴリズムを用いてデータにラベルを付与することでラベル付きデータが少ない場合における半教師あり学習の頑健性を高める手法を提案する。

以下に本論文の構成を示す。まず、2章では本研究で用いる要素技術について概説する。続いて3章で実験手法の提案をし、4章において実験結果と考察を示す。5章で本研究の成果をまとめたうえで、今後の課題について述べる。

2 要素技術

本章では，実験に関連する要素技術について説明する．

2.1 半教師あり学習

半教師あり学習 (Semi Supervised Learning: SSL)^{[2],[3]} は大量のラベルなしデータと少量のラベル付きデータを用いて学習を行う手法である．

2.2 疑似ラベル

疑似ラベル (Pseudo Label)^[4] はあるモデルによって予測されるラベルなしデータに対する暫定的なラベルである．SSL では疑似ラベルを付与したデータをラベル付きデータに混ぜて学習することで各ラベル同士に対する確率分布を粗密なものにする作用があり，正則化^[5]の役割をする．データ数を N ，あるモデルによる予測確率を q_b ，予測確率 p, q に対する Cross Entropy Loss を $H(p, q)$ ，閾値を τ とするとき，(2.2) 式で与えられる．

$$\hat{q}_b = \operatorname{argmax}(q_b) \quad (2.1)$$

$$\frac{1}{N} \sum_{b=1}^N \mathbb{1}(\max(q_b) \geq \tau) H(\hat{q}_b, q_b) \quad (2.2)$$

2.3 FixMatch

FixMatch^[1] は SSL の一つである．図 2.1 に FixMatch の概略図を示す．疑似ラベルと Consistency Regularization の二つの正則化手法を統合させた手法である．

2.3.1 RandAugment

RandAugment^[6] とは，データ拡張手法の一つであり，チューニングするパラメータは N と M の二種類で，複数あるデータ変換操作のうちランダムに

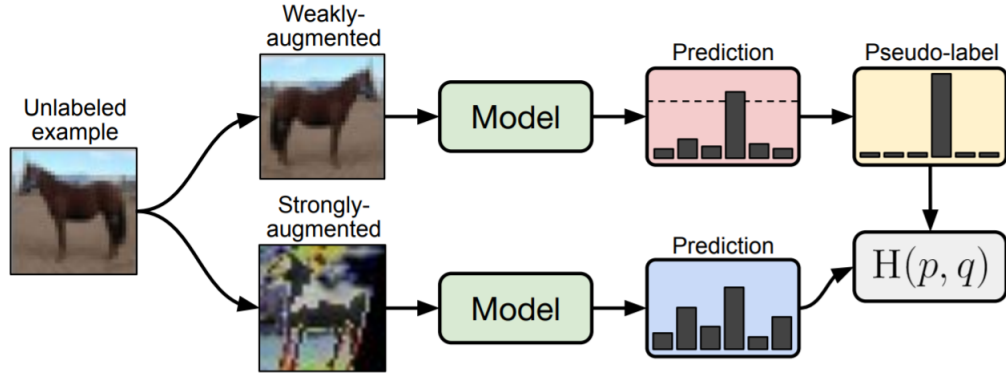


図 2.1: FixMatch の概略図 : (文献^[1]の Figure 1 を参照)

N 個選択し変換強度 M で順に適用する手法である．FixMatch では強変換として用いられる．

2.3.2 Consistency Regularization

Consistency Regularization^[7] は正則化手法の一つであり，画像による変換の前後で予測値が変わらないようなロスを与える．ラベルなしデータを u_b ， $\alpha(\cdot)$ を画像変換， $p_m(y|x)$ を入力 x に対するモデルの出力とすると，(2.3) 式で与えられる．

$$\sum_{b=1}^N ||p_m(y|\alpha_1(u_b)) - p_m(y|\alpha_2(u_b))||_2^2 \quad (2.3)$$

2.3.3 更新

まず，バッチサイズを B としラベル付きデータを $\mathcal{X} = (x_b, p_b) : b \in (1, \dots, B)$ としたとき，ラベル付きデータに対するロスを ℓ_s とすると，(2.4) 式で与えられる．

$$\ell_s = \frac{1}{B} \sum_{b=1}^B H(p_b, p_m(y|\alpha(x_b))) \quad (2.4)$$

また，ラベルなしデータに対するロス ℓ_u は，(2.2)，(2.3) 式より，

$$\ell_u = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbb{1}(\max(q_b) \geq \tau) H(\hat{q}_b, p_m(y|\mathcal{A}(u_b))) \quad (2.5)$$

となる．このとき， μ はバッチ内のラベル付きデータに対するラベルなしデータの比率， $\mathcal{A}(\cdot)$ は強変換である．

従って，ラベルなしデータの重みを λ として，バッチ全体のロス $\ell_s + \lambda \ell_u$ となる．

2.4 SimCLR

Simple framework for Contrastive Learning of visual Representation(SimCLR)^[8] は SSL の一つである．図 2.2 に概略図を示す．モデルの構造は Encoder, Projection Head, Classifier から構成されており，Encoder と Projection Head に対して，全データを用いて Contrastive Learning によって学習する．次に学習済みの Encoder と Classifier に対して，ラベル付きデータを用いて Classifier の学習を行う．最後にモデルの蒸留^[9]を行うことで最終的なモデルを得る．

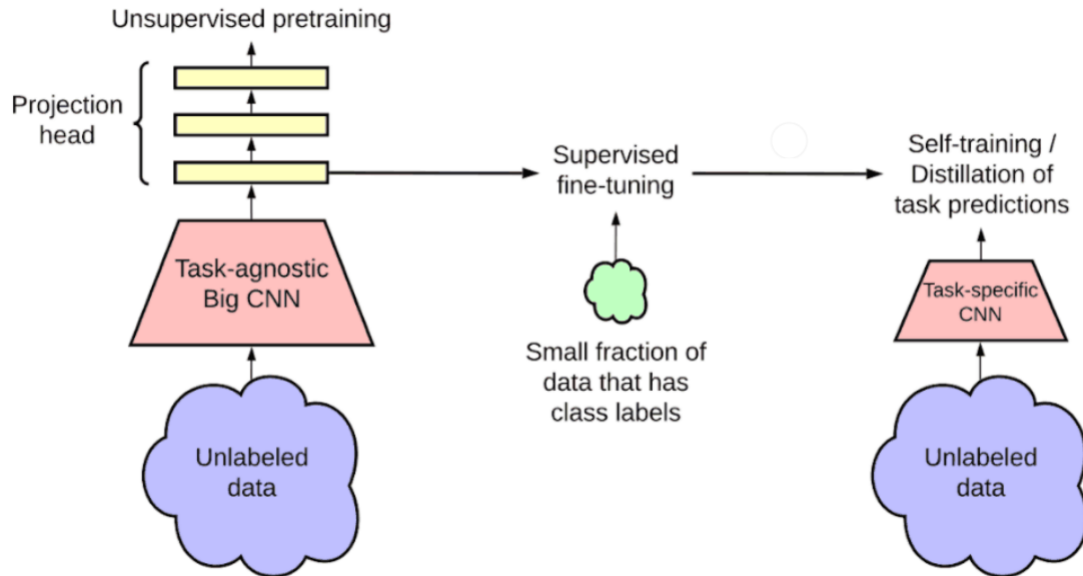


図 2.2: SimCLR の概略図：(文献^[10]の Figure 3 を参照)

2.4.1 Cotrastive Learning

Contrastive Learning (CL)^[11] とは特徴表現を獲得するための自己教師あり学習^[12]のひとつである。ある画像変換をした画像のペアについて元画像が一致するか否かを識別するタスクであり、一つの画像から得られる特徴表現が画像変換によって画像の持つ意味が大きく変化しないような Encoder を獲得することができる。具体的な方法について、バッチ内の画像枚数を N 枚とすると、Data Augmentation によって 2 倍に増やしたとき、各画像に対して正例は 1 枚、負例は $2(N - 1)$ 枚となる。このとき特徴ベクトル間の距離として \cos 類似度を用いると、(2.6) 式で表される。また、正例との距離を小さくかつ、負例との距離を大きくするために正例のペア $(\mathbf{z}_i, \mathbf{z}_j)$ に対するロス $\ell(i, j)$ は (2.7) 式で表され、バッチ全体のロス \mathcal{L} は (2.8) 式で表される。

$$\text{sim}(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T \mathbf{v} / \|\mathbf{u}\| \|\mathbf{v}\| \quad (2.6)$$

$$\ell(i, j) = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)} \quad (2.7)$$

$$\mathcal{L} = \frac{1}{2N} \sum_{k=1}^N [\ell(2k - 1, 2k) + \ell(2k, 2k - 1)] \quad (2.8)$$

2.5 ResNet

Residual Network (: ResNet)^[13] は Deep Neural Network (: DNN)^[14] のモデルの一つであり、DNN において層を深くすることで発生する劣化問題及び勾配消失問題^[15]を解消するために残差についての学習を行うモデルである。図 2.3 に ResNet の構成要素である Residual Block の構造を示す。2 層の畳み込み層とショートカットを足し合わせた構造となっている。

2.6 Genetic Algorithm

遺伝的アルゴリズム (Genetic Algorithm: GA)^[17] とは、生物の進化を模倣した組合せ最適化問題のアルゴリズムである。解の要素の最小単位を遺伝子、

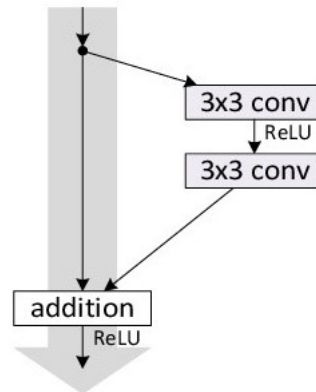


図 2.3: Residual Block の構造 : (文献^[16] の Figure 2.(a) を参照)

遺伝子の集まりである解を個体として表現する。また個体の集合を世代とし、各個体の計算された適応度をもとに選択、交叉、突然変異の3つの操作によって新たな個体群を生成し、次世代の個体集合とする。繰り返し世代を重ねることで最適解を探索する。個体の基本的なエンコーディング方法として、バイナリ型、順列型、実数型、整数型がある。本研究では整数型について扱うため、以下整数型を前提とした説明をする。

2.6.1 選択

選択^[18]は自然淘汰をもとにした操作であり、個体の適応度にもとづき次世代に残される個体を選ぶものである。以下のものがあげられる。

エリート選択^[19]： 世代における適応度の最も高い個体を他の操作を行わず次世代に残す手法である。

トーナメント選択： 個体群からランダムに決められた数 (: トーナメントサイズ) 取り出し、その中で適応度の最も高い個体を選択する手法である。

2.6.2 交叉

交叉は生物の交配をもとにした操作であり，2つの個体から新たな2つの個体を生成するものである．以下のものがあげられる．

二点交叉： 対象となる2つの個体を同じ遺伝子座で3つに分割を行い，いずれかを入れ替える手法である．

一様交叉： 対象となる2つの個体についてある遺伝子座ごとにある確率で入れ替えを行う手法である．

2.6.3 突然変異

個体の遺伝子を変化させる操作で，局所探索になることを防ぐ．乱数によって他の取りうる値に変化させる．また，突然変異率を上げすぎるとランダム探索となり，収束しなくなるため，高くても数%に設定されることが多い．

3 提案手法

本研究では，ラベルなしデータに対する疑似ラベルを GA を用いて探索する手法を提案する．また以降，半教師あり学習のデータについて学習データであるラベル付きデータとラベルなしデータをそれぞれ D_l ， D_{ul} ，テストデータを D_t と呼ぶ．

3.1 個体設定

まず， D_{ul} からランダムにいくつかデータを取り出し探索データとする．ここで，GA の扱う個体は探索データに対する疑似ラベル群である．探索データと各遺伝子座は一対一対応しており，遺伝子型は対応するラベルを表す整数値である．従って，遺伝子長は探索データ数となる．以降探索データについて D_s と呼ぶ．

4 提案手法

以下に提案手法の手順を示す．

1. モデルの事前学習
2. 事前学習したモデルに入力データを D_s ，ラベルデータを各個体とし学習し， D_l の識別率を適応度として GA の探索を行う．
3. D_l と探索された個体をラベルとして持つ D_s とを合わせてラベル付きデータとして SSL によって再学習を行い D_t を識別する．

既存の疑似ラベルを用いた手法では D_s で学習されるモデルの精度に大きく影響されるが，提案手法では疑似ラベルはモデルの出力ではないため正答率が上がり，結果として半教師あり学習による精度も改善されることが期待できる．

5 数値実験

本研究ではデータセットとして CIFAR-10^[20] を用いた。また実験 1, 2 では半教師あり学習のベースモデルとしてそれぞれ FixMatch , SimCLR を用いて GA によるラベル探索及び性能評価を行った。FixMatch と SimCLR で実験を行うことについて、同じ実験設定で同程度の精度を出すことができる一方、FixMatch は毎回モデル全体を学習するのに対し SimCLR は出力層である分類器のみが学習される違いがある。

5.1 実験 1

5.1.1 実験設定

表 5.1, 5.2, 5.3 にそれぞれ実験 1 における各種設定を示す。また、FixMatch は学習に時間がかかるためラベル付きデータの一部を用いて事前学習をしたモデルの重みを初期重みとして使用する。

表 5.1: 実験 1 : データ内訳

D_l	250
D_{ul}	49650
D_s	100
D_t	10000

表 5.2: 実験 1 : GA の設定

個体数	20
世代数	20
選択	エリート + トーナメント
エリート数	2
トーナメントサイズ	2
交叉	二点交叉
交叉率	1.0
突然変異	ランダム遷移
遺伝子座ごとの突然変異率	0.02

表 5.3: 実験 1 : FixMatch の設定

model	WideResNet16-2	
data set	cifar10	
batch size	labeled	32
	unlabeled	$32 * 7$
optimizer	SGD(lr=0.1,momntum=0.9)	
loss	cross_entropy_loss	
事前学習		
train	labeled	$100(:D_l)$
	unlabeled	$49650(:D_{ul})$
val data	150	
num_iterations	2^{15}	
GA の評価		
train	labeled	$100(:D_s)$
	unlabeled	$49650(:D_{ul})$
validation data	$250(:D_l)$	
num_iterations	5000	
探索された個体の評価		
train	labeled	$250(:D_l)+D_s$
	unlabeled	$49650(:D_{ul})$
test data	$10000(:D_t)$	
num_iterations	2^{16}	

5.1.2 結果と考察

図 5.1, 5.2 に GA の探索結果を示す．図 5.1 は横軸が GA の世代，縦軸は箱ひげ図が適応度，折れ線グラフは疑似ラベルの正答率で，図 5.2 は横軸が疑似ラベルの正答率で縦軸が適応度を示している．図 5.2 から正の弱い相関性があることが確認できる一方で，図 5.1 では適応度のばらつきは減っており，学習自体は行われていることが確認できるものの疑似ラベルの正答率は全く向上していない．このことから，疑似ラベルの正答率と識別率は関係性があるが，誤ラベルの中に少量のラベル付きデータに対し過剰に適合するような局所解となりうるものの存在があると考えられる．

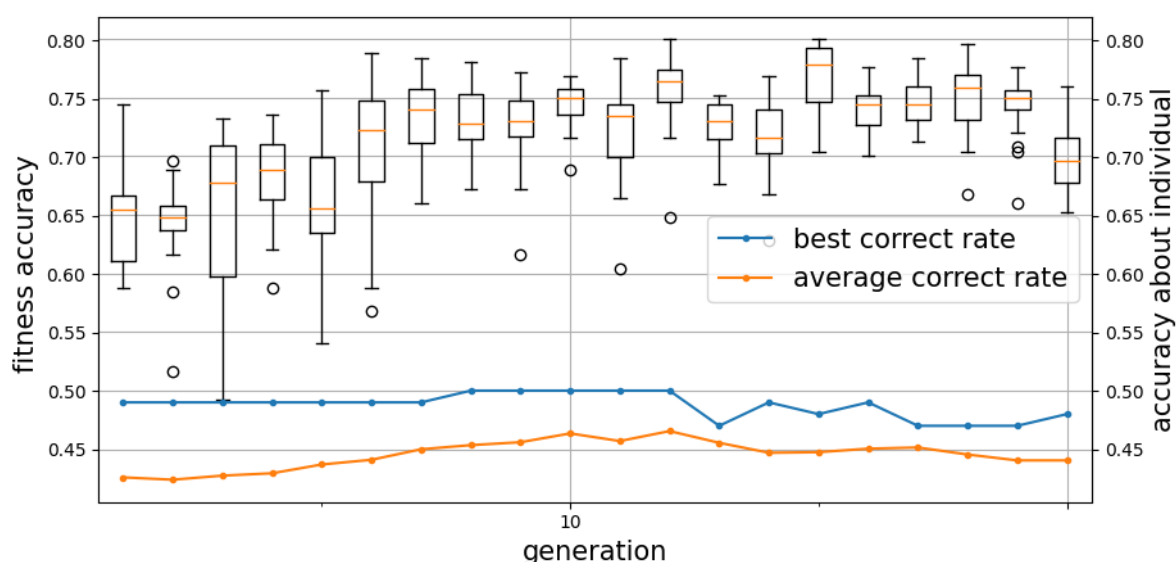


図 5.1: 実験 1 : GA の探索結果

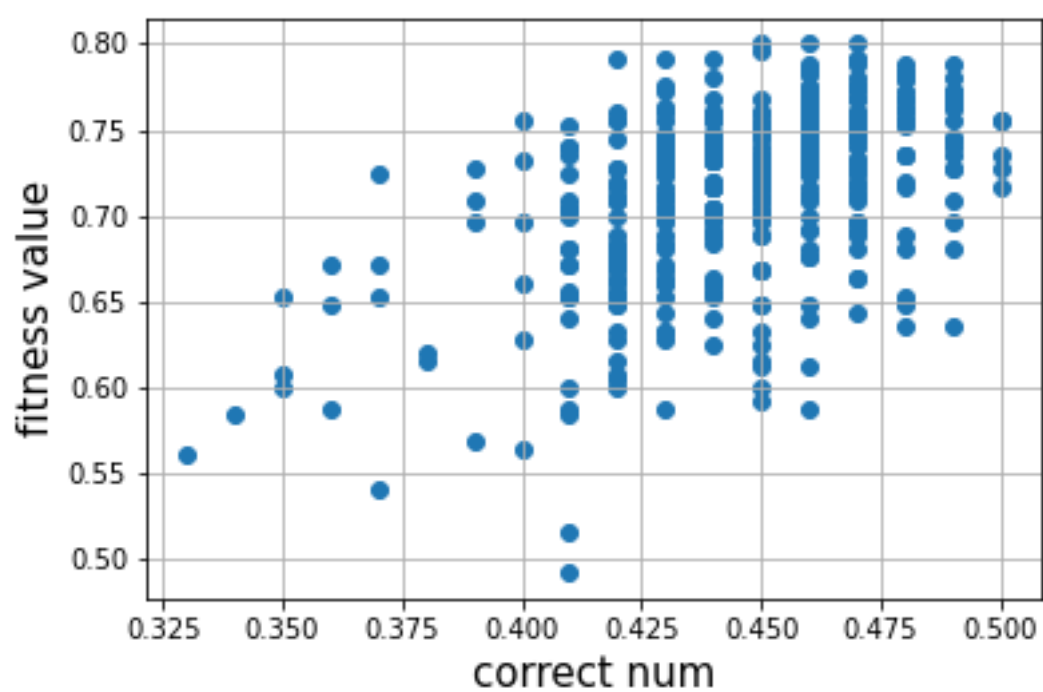


図 5.2: 実験 1 : 探索個体の散布図

次に、表 5.4 に探索された個体によるテスト識別率を示す。図 5.1 より適応度の平均が最大であった 15 世代の全個体のうち各遺伝子座で最も多く選ばれた遺伝子を採択して得られる新たな個体を最終的な個体として学習を行った。またある遺伝子座について選ばれた遺伝子が 14 世代の全個体に対して占めた割合に対し確信度として閾値を設けて、採択数を絞った状態での学習も行った。

結果は全て探索した疑似ラベルを使用しないベースラインを下回った。考えられることとして、まず疑似ラベルの精度が低いことによって疑似ラベルによる正則化がうまく機能しなかったことが挙げられる。また、FixMatch のラベルなしデータによる疑似ラベルの正則化と、ラベル付きデータとして混ぜられた疑似ラベルの正則化が互いに妨害をした可能性が考えられる。

表 5.4: 実験 1 : テスト識別率

閾値	採択数	正答数	テスト識別率
	0		0.868
なし	100	46	0.836
0.19	39	22	0.862
0.2	19	14	0.825

5.2 実験 2

表 5.5, 5.6, 5.7 にそれぞれ実験 2 における各種設定を示す。また、事前学習は SimCLR の Encoder 部のみとし、Classifier の初期重みはランダムなものとした。

表 5.5: 実験 2 : データ内訳

D_l	50
D_{ul}	49850
D_s	100
D_t	10000

表 5.6: 実験 2 : GA の設定

個体数	50
世代数	100
選択	エリート + トーナメント
エリート数	2
トーナメントサイズ	2
交叉	一様交叉
個体の交叉率	1.0
遺伝子座ごとの交叉率	0.5
突然変異	ランダム遷移
遺伝子座ごとの突然変異率	0.1

表 5.7: 実験 2 : SimCLR の設定

model	Encoder	ResNet18
	Projection head	2 層 MLP(shape:2048to512)
	classifier	MLP(shape:2048to10)
事前学習		
train data	unlabeled	50000($:D_l + D_{ul} + D_s$)
batch size	1024	
epochs	500	
optimizer	RAdam(lr= $1.0 * 10^{-3}$)	
GA の評価		
train data	labeled	100($:D_s$)
batch size	16	
epochs	25	
loss	Cross Entropy Loss	
optimizer	Adam(lr= $5.0 * 10^{-3}$,momntum= $1.0 * 10^{-6}$)	
validation data	50($:D_l$)	
探索された個体の評価		
train data	labeled	150($:D_l + D_s$)
batch size	16	
epochs	100	
loss	Cross Entropy Loss	
optimizer	Adam(lr= $1.0 * 10^{-3}$,momntum= $1.0 * 10^{-6}$)	
test data	10000($:D_t$)	

5.2.1 結果と考察

図 5.3, 5.4 に GA の探索結果を示す．図 5.3 は横軸が GA の世代，縦軸は箱ひげ図が適応度，折れ線グラフは疑似ラベルの正答率で，図 5.4 は横軸が疑似ラベルの正答率で縦軸が適応度を示している．図 5.4 から正の強い相関性があることが確認でき相関係数は 0.822 である．また，図 5.3 では適応度，疑似ラベルの正答率ともに上昇していることから，当然ではあるが，ラベル付きデータの精度と学習されたモデルの識別率は強い相関関係があることが分かる．一方で，疑似ラベルの精度は 0.37 と低い値に収まっており，探索空間が膨大なために局所解に陥った可能性が考えられる．

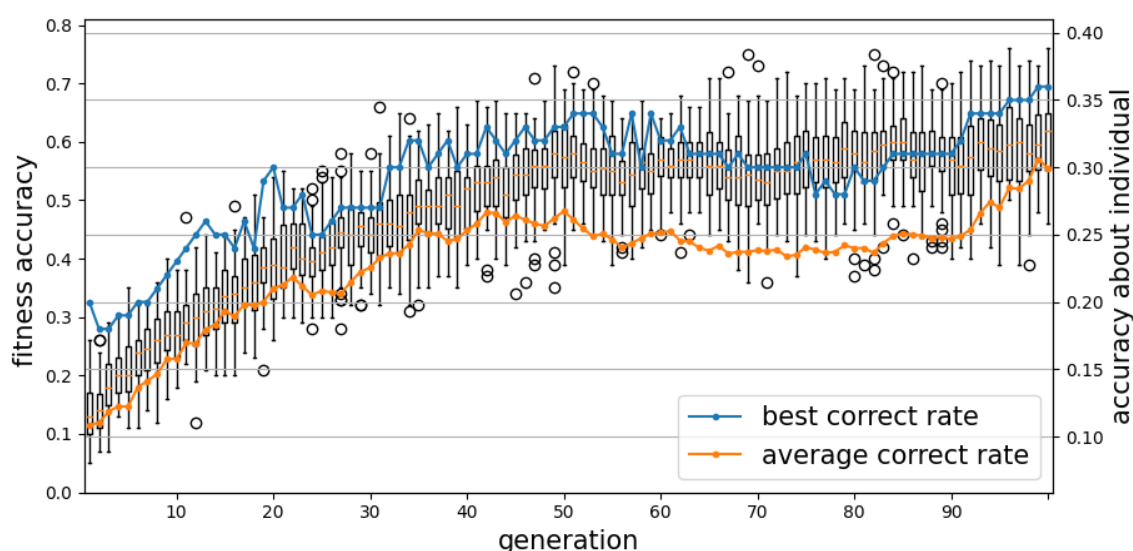


図 5.3: 実験 1 : GA の探索結果

次に，表 5.8 に探索された個体によるテスト識別率を示す．結果は提案手法による疑似ラベルがベースラインおよび従来の疑似ラベルよりもテスト識別率が劣っていることがわかる．疑似ラベルの精度が低いためテスト識別率の精度も低くなっていると考えられる．

また，実験 2 では Encoder の学習時間が 1days GPU かかってはいるものの，1 個体の適応度を出す時間について，実験 1 に対し 16 倍ほど速くすることができ，個体数や世代数を多くするにはベースモデルとして FixMatch よりも SimCLR を用いる方が良いことが分かる．

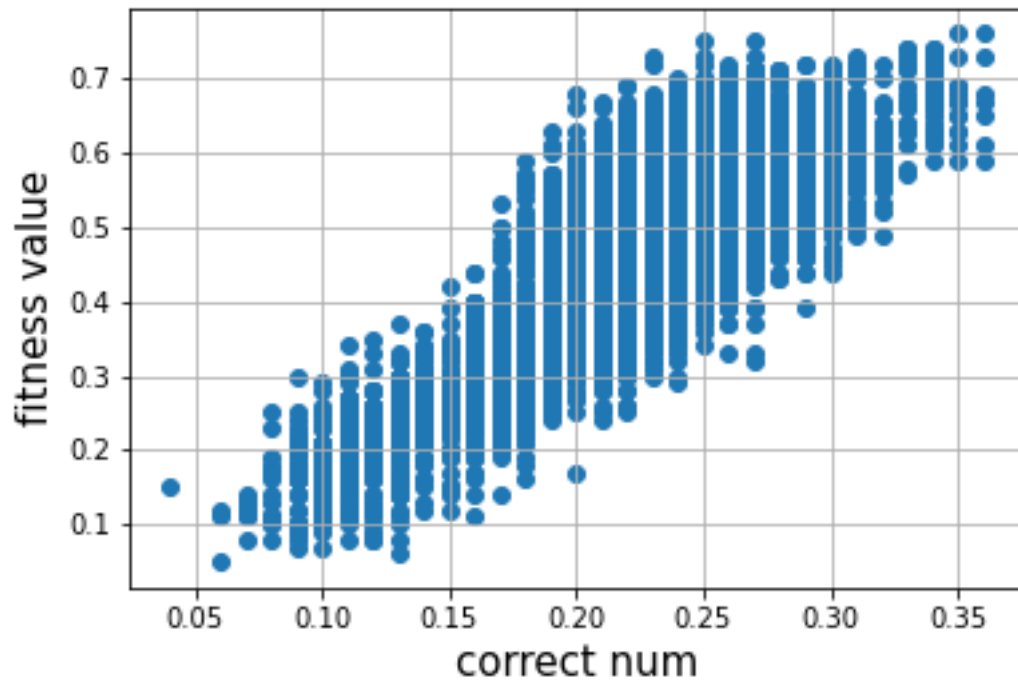


図 5.4: 実験 2 : 探索個体の散布図

表 5.8: 実験 2 : テスト識別率

D_s に対するラベル	テスト識別率	正答率
正解ラベル	0.822	1.00
提案手法による疑似ラベル	0.674	0.37
baseline モデルによる疑似ラベル	0.784	0.74
baseline (ラベル付きデータのみ)	0.772	\

6 まとめと今後の課題

本研究では，遺伝的アルゴリズムを用いてラベルなしデータに対する疑似ラベル生成手法の提案をした．また，数値実験から探索された疑似ラベルを用いた学習において精度の低下を確認し，遺伝的アルゴリズムによるラベル探索は非常に困難であることが分かった．

今後の課題として，遺伝的アルゴリズムのパラメータチューニング及び，膨大な探索空間に有効に拡張されたアルゴリズムを使用するといった遺伝的アルゴリズムの改善と，validation data についても画像の水増しを行うことや一個体の学習に対するパラメータチューニングといった適応度計算の改善の2点が挙げられる．

謝辞

本研究を進めるにあたりご指導，ご鞭撻を賜りました森直樹教授，岡田真助教授に深く感謝申し上げます。また，発表スライドや本論文の作成にあたり添削及びアドバイスを下さったソフトウェアシステム研究グループの皆様にも厚く御礼を申し上げ，感謝の意を表します。

2021 年 2 月 26 日

参考文献

- [1] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020.
- [2] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. 2005.
- [3] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.
- [4] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, 2013.
- [5] Yves Grandvalet and Yoshua Bengio. Entropy regularization., 2006.
- [6] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020.
- [7] Han Zhang, Zizhao Zhang, Augustus Odena, and Honglak Lee. Consistency regularization for generative adversarial networks. *arXiv preprint arXiv:1910.12027*, 2019.
- [8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [9] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

-
- [10] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*, 2020.
 - [11] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. What makes for good views for contrastive learning. *arXiv preprint arXiv:2005.10243*, 2020.
 - [12] Carl Doersch and Andrew Zisserman. Multi-task self-supervised visual learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2051–2060, 2017.
 - [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
 - [14] Hugo Larochelle, Yoshua Bengio, Jérôme Louradour, and Pascal Lamblin. Exploring strategies for training deep neural networks. *Journal of machine learning research*, 10(1), 2009.
 - [15] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116, 1998.
 - [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
 - [17] Darrell Whitley. A genetic algorithm tutorial. *Statistics and computing*, 4(2):65–85, 1994.
 - [18] Tobias Blickle and Lothar Thiele. A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4):361–394, 1996.

-
- [19] Tadahiko Murata, Hisao Ishibuchi, and Hideo Tanaka. Multi-objective genetic algorithm and its applications to flowshop scheduling. *Computers & industrial engineering*, 30(4):957–968, 1996.
- [20] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.