

jh190066-DAH

Deep Neural Network Optimization Based on Dual Inheritance Theory and Its Application

篠崎 隆宏 (東京工業大学)

概要

深層ニューラルネットワークの構成および学習条件を指定するハイパーパラメタの最適化法として、進化戦略と知識蒸留を組み合わせる二重相統進化戦略手法を提案した。従来の進化戦略ではエラー率の情報のみが次世代へ受け継がれるのに対し、知識蒸留を組み合わせることにより世代間で伝えられる情報量を多くすることが出来る。この手法を End-to-End 型の音声認識システムに適用し、従来法より精度が高くコンパクトなモデルが作成できることを示した。

1. 共同研究に関する情報

(1) 共同研究を実施した拠点名

京都大学学術情報メディアセンター

(2) 共同研究分野

超大規模数値計算系応用分野

超大規模データ処理系応用分野

超大容量ネットワーク技術分野

超大規模情報システム関連研究分野

(3) 参加研究者の役割分担

南條 浩輝 (京都大学准教授)

言語科学や比較言語学とともに情報処理に詳しい。二重継承理論を深層ニューラルネットワークを用いた音声認識等の情報処理システムの自動最適化に工学応用する本研究において、アルゴリズム設計および実験結果解析に協力する。

2. 研究の目的と意義

本研究の目的は、二重継承理論と大規模並列計算に基づき、ニューラルネットワークの構造とパラメタを同時に効率的に最適化する自動開発フレームワークを実現することである。

ニューラルネットワークを応用することで知的情報処理の様々な既存タスクで従来

を上回る性能が実現し、また従来不可能であった柔軟な情報処理が実現しつつある。しかし、ニューラルネットワークを用いたシステムを設計し実際に高い性能を実現するためには、ネットワーク構造や学習条件の高度なチューニングが必要となる。現在チューニングは専門家の労力とノウハウに頼っており、システム開発のボトルネックになっている。

我々はこれまでに、深層ニューラルネットワークの構造や学習条件のチューニングを進化計算により自動化する研究を行い、音声認識システムや機械翻訳システムに応用する先駆的な研究を行ってきた。これは、深層ニューラルネットワークの構造や学習条件を埋め込み表現した遺伝子の確率分布を遺伝子のサンプリングと評価を繰り返すことで数値最適化するものである。本技術により Tsubame 上で最適化を行った高性能日本語音声認識システムが世界的に用いられている音声認識ツールキット Kaldi に採用されるなどの成果を得ている。しかし、進化計算による大規模深層ニューラルネットワークのチューニングは計算効率が悪く、また最終的に達成される性能向上にも限界があった。

深層ニューラルネットワークを進化計算において最適化する際、ネットワーク構造や学習条件の最適化と、それに基づいたニューラルネ

ネットワークの結合重みの学習という、最適化の二重構造が存在している。進化最適化では遺伝子が表現するネットワーク構造や学習条件の性能を、深層ニューラルネットのパラメタを学習し評価することを通したシステムの性能で評価している。その際、システムの性能を示す数値のみが遺伝子分布の更新に利用され、学習したニューラルネットのパラメタはすべて捨てられている。

他方、進化生物学や文化人類学分野において人類の知性が他の生物と比較して際立っていることを説明する理論として、二重継承理論がある。人間は親世代における知識を文化として子孫に継承しているが、これは脳のデザインを含む遺伝子継承と並行して行われる。これにより知識の蓄積が可能になるとともに、生物学的な進化と文化的な進化の間に相互作用が発生し、人類の知能と文化の急速な進化につながったと考えられている。

本研究では深層ニューラルネットの進化最適化において、二重継承理論にならない親世代の深層ニューラルネットが学習したネットワークパラメタを次世代で有効活用する方法として、進化戦略アルゴリズムに機械学習における Teacher-student 学習（あるいは知識蒸留）手法を導入する手法を提案し研究

する。

3. 当拠点公募型研究として実施した意義
本研究では進化計算の枠組みのもと大規模な深層ニューラルネットの学習を多数繰り返す。多数の GPU 計算ノードを用いた並列計算を活用する研究であり、高度な HPC 環境が提供されるとともに大学間での連携が行える本拠点公募型研究は当研究を進める上で有用であった。
4. 前年度までに得られた研究成果の概要
当プロジェクトは 1 年間のプロジェクト（2019 年度）のプロジェクトであり、本項は該当しない。
5. 今年度の研究成果の詳細
A)
従来の進化戦略による深層ニューラルネットの最適化法では個体の評価スコアのみが次世代に伝えられる。本研究では生物学における二重相続理論をもとに、図 1 に示すように遺伝子と並行して祖先世代の知識が次世代に伝達される 2 重相続進化戦略を提案し

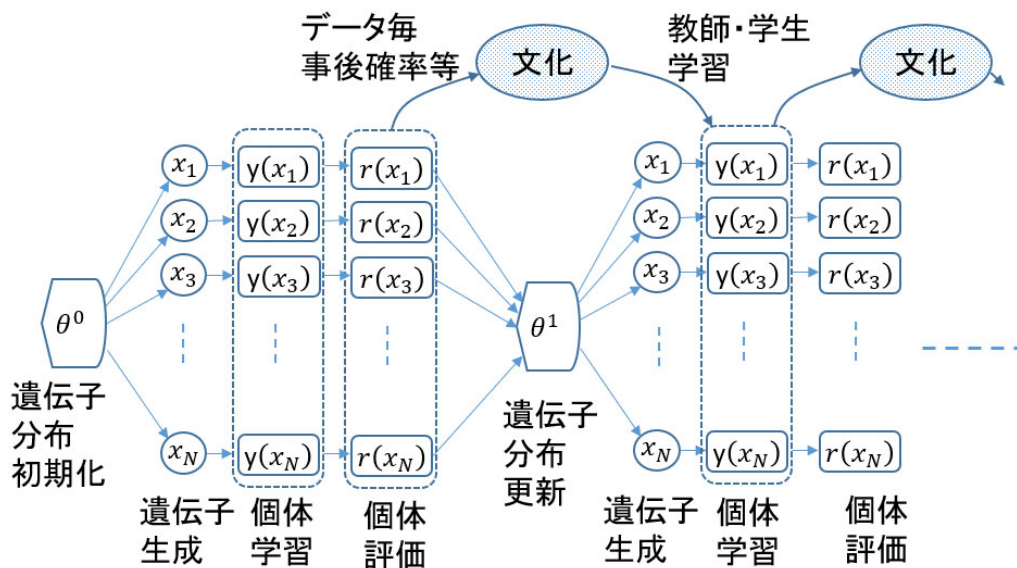


図 1 提案 2 重相続進化戦略

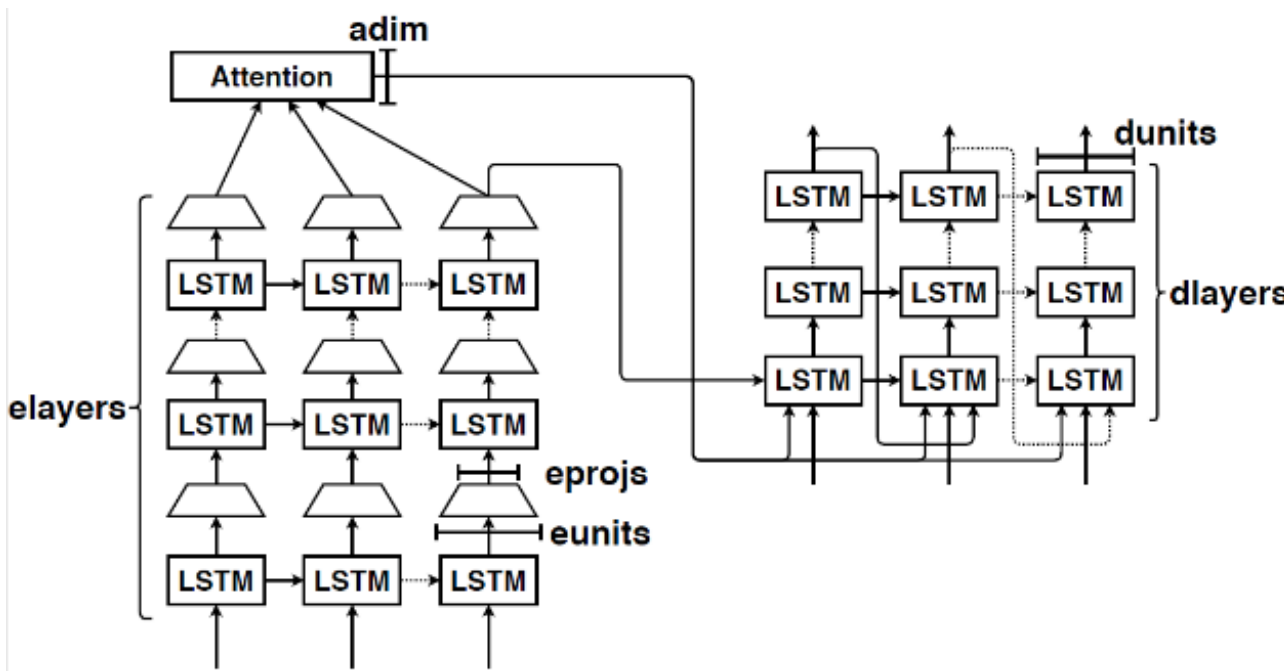


図 2 End-to-End 音声認識システムのネットワーク構成

た。世代とともに蓄積される知識は文化ととらえることができる。深層ニューラルネットの学習において、祖先世代の個体から子孫世代の個体へ何を伝えるかは任意である。本研究では、深層学習研究において広く用いられている知識蒸留の手法を個体間で知識を伝達する方法として用いることを検討した。

知識蒸留は、学習済みの深層ニューラルネットを用いて学習データに対してカテゴリ予測を行い、カテゴリ予測の事後分布をソフトターゲットとして別の深層ニューラルネットの学習に用いるものである。学習済みのニューラルネットを教師ネットワーク、学習を行うニューラルネットを生徒ネットワークと呼ぶ。教師ネットワークによる学習データに対する解釈を生徒ネットワークに伝えていると解釈できる。

実験では、End-to-End 音声認識システムを最適化対象の深層ニューラルネットとして用いた。提案法のアイデアは様々な進化的アルゴリズムに適用することが可能であるが、本研究では効率的な進化戦略法として知ら

れている共分散行列適応進化戦略をベースの進化アルゴリズムとして用いた。また、組み込む知識蒸留の手法として複数の手法を実装して比較した。

B)

ハイパーパラメタの最適化手法の一種に共分散行列適応進化戦略(CMA-ES)がある。この手法では最適化したいハイパーパラメタ群を遺伝子とし、遺伝子の確率分布を多変量ガウス分布で表す。ガウス分布からサンプリングにより遺伝子集合を求め、各遺伝子に対応したハイパーパラメタで個体の評価スコアを計算し、ガウス分布を更新する。このプロセスを1世代とし、世代数を重ねることでパラメタの最適化を行う。遺伝子の値に対してシステムのハイパーパラメタ設定が決まる。設定されたハイパーパラメタに対してシステムの性能が決まるが、多くの場合ハイパーパラメタと性能の関係は複雑であり関数として陽に表現することができずブラックボックス最適化の問題となる。CMA-ESは遺伝子

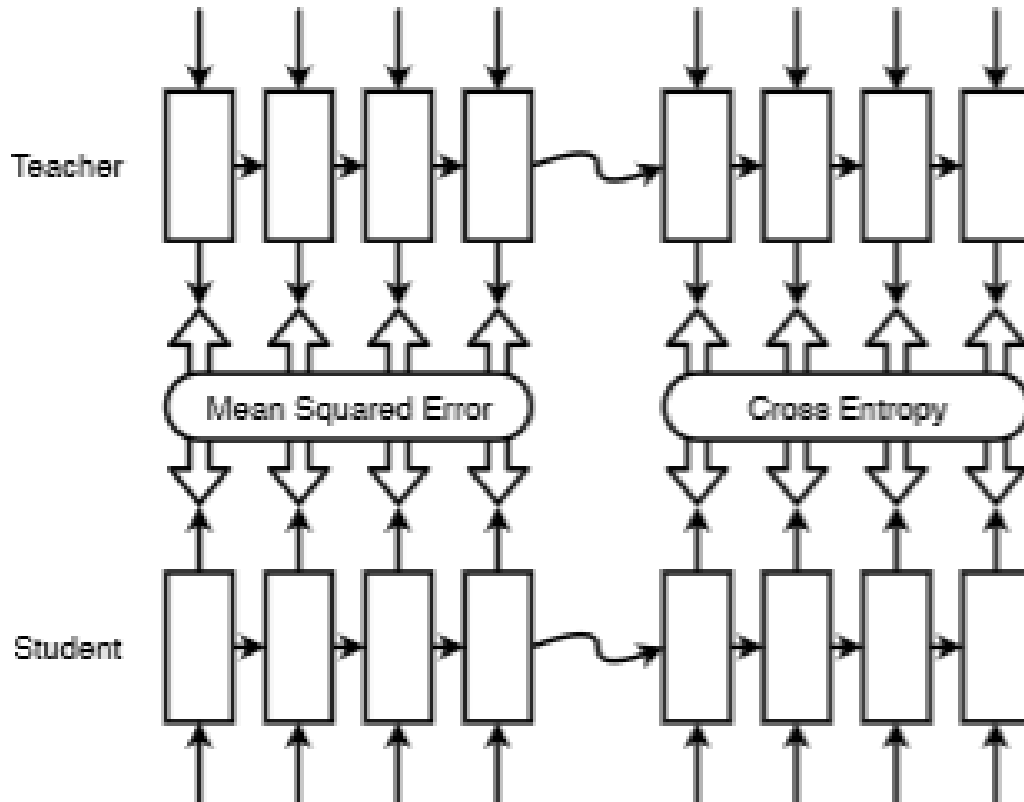


図 3 End-to-End ネットワークの知識蒸留

に対応するシステムの性能を評価することが出来さえすれば、ハイパーパラメタの最適化を行うことができる。

C)

実験には End-to-End 音声認識システムのツールキットである ESPnet を用いた。ESPNet には、各種のデータセットに対して音声認識システムの構築と評価を行うスクリプト群が用意されている。本研究ではその中で、an4 レシピを用いた。an4 の学習セットの話者は 74 名である。948 の発話が記録され、平均 3 秒、合計約 50 分の音声記録されている。評価セットの話者は 10 名である。130 の発話が記録され、合計は約 6 分の音声記録されている。

図 2 に使用した ESPnet の End-to-End 深層ニューラルネットの構造を示す。ネットワークは音声特徴量系列を入力として内部表現

を得る Encoder ネットワークと、内部表現から出力音声認識結果を生成する Decoder ネットワークから構成されている。また、図には示していないが、CTC ネットワークも併用する仕組みになっている。

D)

図 3 に、End-to-End 深層ニューラルネットワークへの知識蒸留の組み込み方法を示す。図で上側が教師ネットワーク、下側が生徒ネットワークである。また左側が Encoder ネットワーク、右側が Decoder ネットワークである。

基本となるのは、ネットワーク出力において認識対象のカテゴリの事後確率を教師ネットワークから生徒ネットワークに伝える方法である。この場合、クロスエントロピーによる誤差尺度を用いた。また、拡張版として、Encoder が出力する内部表現出力を教師

表 1 最適化対象のハイパーパラメタ

種類	ハイパーパラメタ	初期値
一般	patience	3
	mtlalpha	0.5
エンコーダ	elayers	4
	eunits	320
	eprojs	320
デコーダ	dlayers	1
	dunits	300
Attentions	adim	320
	aconv-chans	10
	aconv-filts	100
知識蒸留	μ	0.3

ネットワークから生徒ネットワークに伝達させることも試みた。この場合はベクトル間の距離尺度として平均 2 乗誤差を誤差尺度に用いた。

実装は、生徒となる End-to-End 深層ニューラルネットのクラス内に教師役のニューラルネットの構造を追加し、バックプロパゲーションにおける前向き計算を同時に行ったうえで、教師ネットワークにおける後ろ向き計算をブロックする形で行った。

E)

表 1 に最適化の対象としたハイパーパラメタを示す。ハイパーパラメタには、ネットワークの学習条件に関するパラメタ、ネットワークの構造に関するパラメタが含まれている。また知識蒸留を行う際に、祖先からの知識と学習ラベルのどちらに重点を置いて学習するか、の重みパラメタがある。祖先からの知識みに頼るのでは新しい学習ができず、学習ラベ

ルのみに頼るのでは祖先から子孫への知識の伝達が行えない。本研究ではこの重みも、進化における最適化対象の一つとして含めた。評価実験では比較として従来の CMA-ES による進化も行うが、知識蒸留の重みパラメタは提案法のみで用いる。

進化の起点となる初期個体には、an4 レシピの設定を用いた。提案法における教師個体は、全祖先世代の個体の中から初期個体よりもサイズの大きい個体のなかで認識エラー率の最も小さいものを選択した。第一世代の教師は初期個体とした。

進化の目的関数として、ネットワークサイズと認識誤り率の重み和を用いた。ネットワークの進化実験は CMA-ES と DI-ES のそれぞれについて、個体数 15, 25, 50 の 3 種類の条件で行った。各個体の最大 epoch 数は 40 であり、第 15 世代まで進化を進めた。

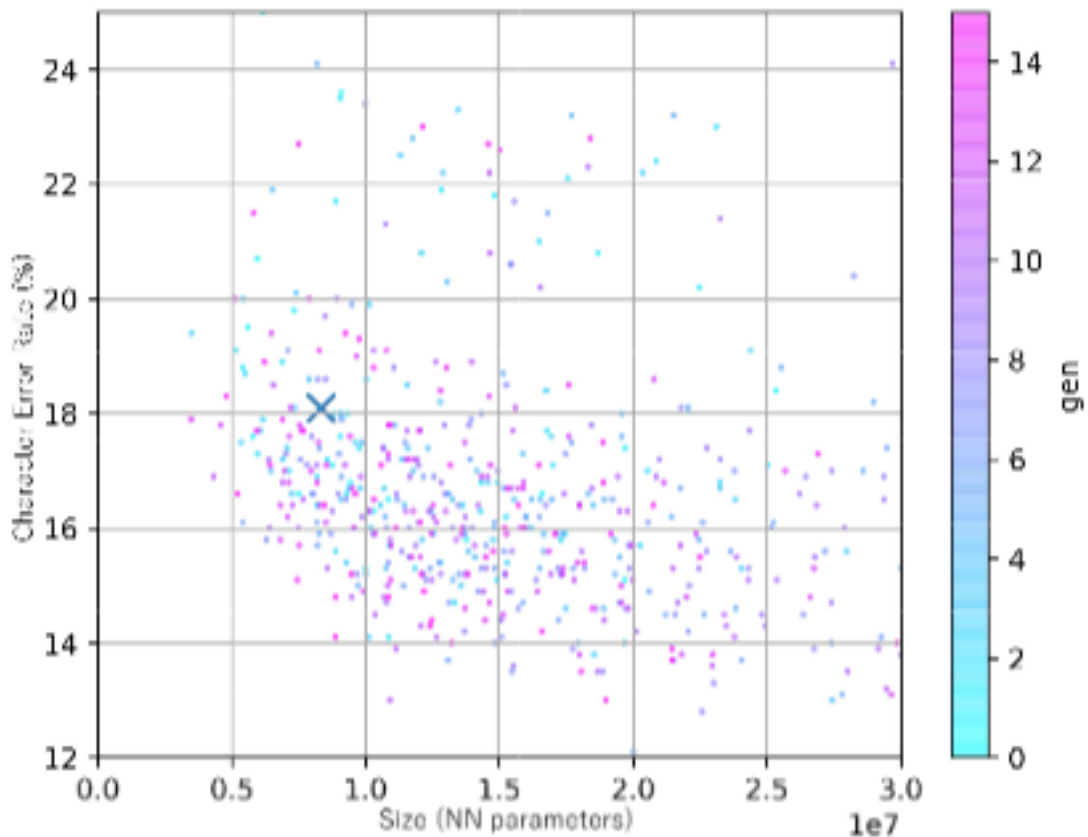


図 4 CMA-ES による進化例 (×印は初期個体)

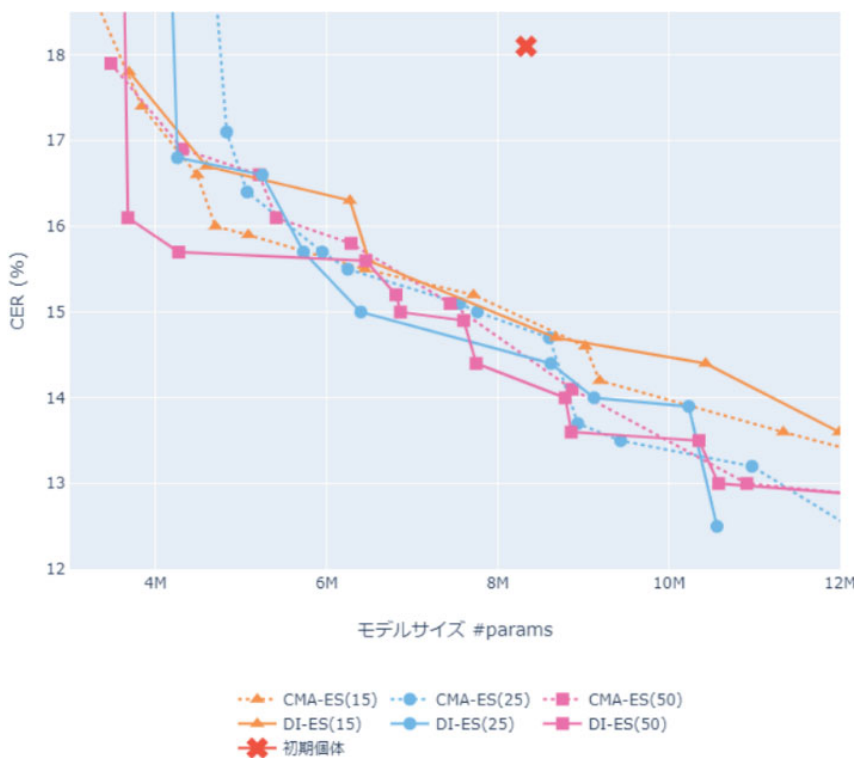


図 5 LOSS_MSE を用いた場合の進化結果 (パレートフロンティアを表示)

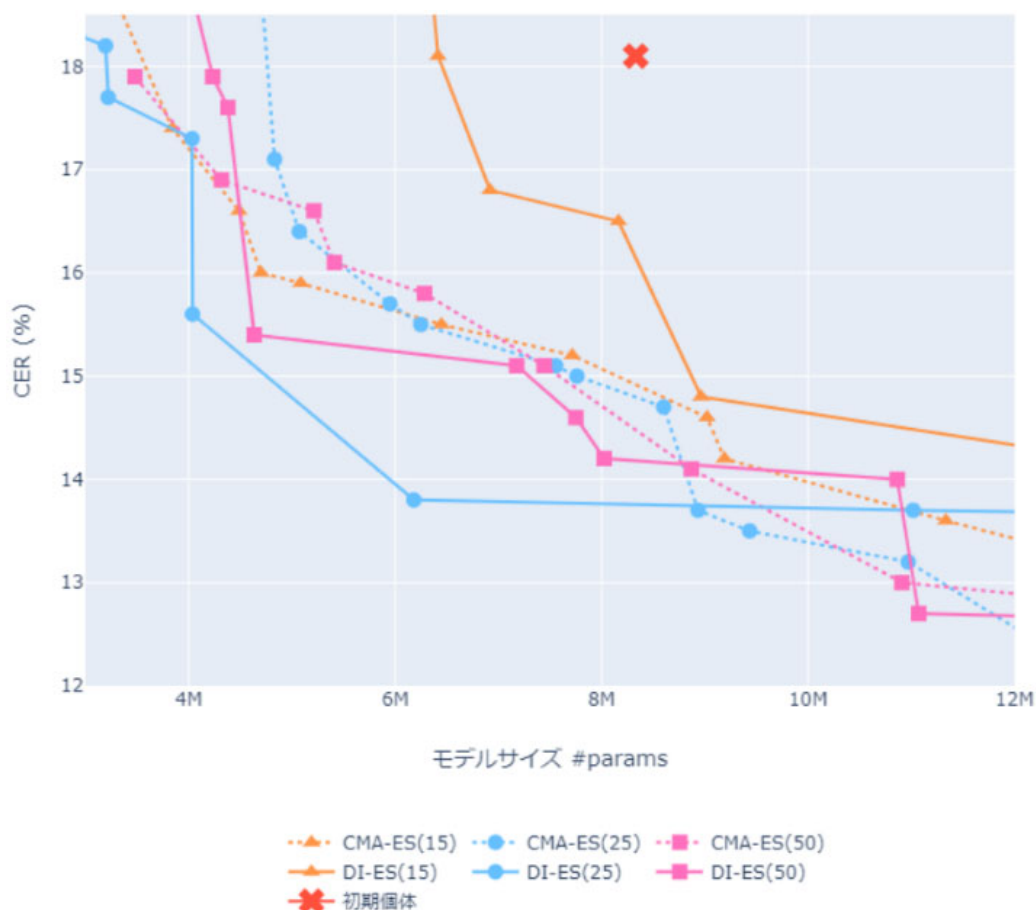


図 6 LOSS_CE を用いた場合の進化結果 (パレートフロンティアを表示)

F)

図 4 に、ベースラインとなる CMA-ES を用いた場合の進化結果を示す。横軸がネットワークのサイズ、縦軸が認識誤り率である。サイズが小さいほど計算量の点で有利であり、エラー率が低いほど認識性能が優れている。×印で示しているのが、初期個体である。世代が進むごとに、サイズが小さくエラー率の低い個体が多く出現していることがわかる。15 回の進化ののち、初期個体よりもモデルサイズとエラー率の両方で初期個体よりも大きく優れた個体が見られることがわかる。

図 5 および 6 に、従来 CMA-ES と提案法による進化実験の結果を示す。横軸と縦軸の解釈は図 4 と同じであるが、パレートフロンテ

ィアのみを表示している。いずれの進化手法・条件においても、ESPnet の初期個体よりも大幅に優れた結果が見られている。従来法と提案法を比較すると、従来法は個体数を増やしても結果に大きな変化は見られないが、提案法では個体数を増やすと従来法よりサイズが小さくエラー率も低い個体が増加した。特に decoder の出力同士の Loss_CE を追加したものはサイズの小さな範囲でその傾向がみられ、提案法の有効性が確認できた [1-3]。

6. 今年度の進捗状況と今後の展望

本プロジェクトは 1 年間の計画であり、プロジェクトとしては昨年度で終了している。提案する二重相続進化戦略手法の研究は継続

しており、現在は Transformer を用いたより大規模な音声認識システムへ適用する実験を行っている。

7. 研究業績一覧（発表予定も含む）

(1) 学術論文（査読あり）

(2) 国際会議プロシーディングス（査読あり）

[1] Kent Hino, Yusuke Kimura, Yue Dong and Takahiro Shinozaki, "Dual Inheritance Evolution Strategy for Deep Neural Network Optimization," Proc. IEEE Congress on Evolution Computation (CEC) 2020, accepted.

(3) 国際会議発表（査読なし）

(4) 国内会議発表（査読なし）

[2] 日野 健人, 木村 友祐, Dong Yue, 篠崎 隆宏, 「二重相続進化戦略による音声認識システムの最適化」, 日本音響学会 2020 年 春季研究発表会 3 月 16 日—18 日

[3] 木村 友祐, 日野 健人, 董 越, 篠崎 隆宏, 「二重相続進化戦略による End-to-End 音声認識システムの最適化」 第 131 回音声言語情報処理研究会 (SIG-SLP) 2020 年 2 月 13 日(木)・14 日(金)

(5) その他(特許, プレスリリース, 著書等)

[4] Takahiro Shinozaki, Shinji Watanabe, Kevin Duh, "Automated Development of DNN Based Spoken Language Systems Using Evolutionary Algorithms," Book Chapter in Nasimul Noman, Hitoshi Iba, "Deep Neural Evolution - Deep Learning with Evolutionary Computation," Springer, 2020.