

深層学習に有用な隠れ素子数の自動決定法 D-DALP の提案

Proposal of D-DALP: A Method to Automatically Tune the Number of Hidden Units Effective for Deep Learning

山田航佑^{1,2} 大澤正彦¹ 今井倫太¹

Kosuke Yamada^{1,2}, Masahiko Osawa¹, and Michita Imai¹

¹慶應義塾大学

¹Keio University

²カリフォルニア大学マーセド校

² University of California, Merced

Abstract: Although deep learning is an effective method to conduct a machine learning, there is no definite answer on how to initialize the hyper parameters. In this paper, we propose an extension of DALP algorithm, an algorithm to automatically tune the number of hidden units for RBM, Double DALP (D-DALP). D-DALP is an algorithm that automatically tunes the number of hidden units that is effective for deep learning by applying DALP twice to the same data set. In the experiment, it is shown that RBM initialized with D-DALP has a higher identification accuracy when compared to RBM initialized with DALP in a deep learning setting.

1. はじめに

深層学習においてハイパーパラメータの決定は重要な課題である。そこで本研究では入力データセットに柔軟に対応できる隠れ素子数自動決定法について扱う。

著者らは深層学習の構成要素の 1 つとして知られる Restricted Boltzmann Machine (RBM) の隠れ素子数自動決定法を提案している[1]。[1]は少ない隠れ素子を有する複数の RBM を学習させた結果を利用して得られた学習精度を動的に分析および線形化する手法であることから、本論文では Dynamically Analyzed Linear Prediction (DALP) 法と呼ぶ。DALP 法で算出される隠れ素子数は、データセットを学習するために最低限必要な素子数であり、結果として識別タスクで評価した場合に比較的小さい Akaike Information Criterion (AIC) を示す。DALP 法を採用しているモデルとして、著者らが提案した Incremental Learning RBM (IL-RBM) がある[2]。IL-RBM は DALP 法を未学習データに対して繰り返し使用し、ネットワークを拡張することによって RBM での逐次学習を可能にしたモデルである。

しかし、DALP 法の欠点として、AIC の最小化を優先的に行ってしまうため極めて高い精度はあまり見込めない点と、比較的少数の隠れ素子数を算出してしまいうため階層化にはあまり向いていない点が挙げられる。また、IL-RBM は逐次学習を行った際に検出

された未学習データに対して素子を追加する手法であり、バッチ学習を行う際は適用できないほか、DALP 法を採用しているので多層化や高い識別精度を算出することが難しい。したがって、比較的多数の隠れ素子数を算出する自動決定法が必要となる。

本論文は DALP 法を拡張することで、データセット毎に高い学習精度を実現するのに適切な隠れ素子数を算出する Double DALP (D-DALP) 法を提案する。D-DALP 法は、DALP 法によって AIC が最小限になるように構築され、学習を終えたネットワークに再度 DALP 法を適用することによって、多層化による識別精度の向上が望める隠れ素子数を算出する手法である。

本論文の構成は以下のとおりである。第 2 章では RBM 及び DALP 法の基本的な説明を行う。続いて第 3 章では提案手法である D-DALP 法の説明を行う。第 4 章では性能評価として 2 つの実画像データセットに対し DALP 法と D-DALP 法を適用し、その際に獲得する特徴の差に伴う精度の差を検討する。第 5 章に結論を述べる。

2. Restricted Boltzmann Machine

本章では本論文中に用いる Restricted Boltzmann Machine (RBM) およびその拡張モデルである IL-RBM の詳細について説明する。

2.1 RBM の概要

RBM とは、可視層間および隠れ層間の結合が存在

Algorithm 1: DALP 法と D-DALP 法

Function DALP (H)

$m \leftarrow H$ の要素数
for $p = 1$ to m **do**
 $C_p \leftarrow$ 隠れ素子数 H_p で学習済み RBM の
 クロスエントロピー
end for
 $slope \leftarrow \frac{m \sum_{p=1}^m H_p C_p - \sum_{p=1}^m H_p \sum_{p=1}^m C_p}{m \sum_{p=1}^m H_p^2 - (\sum_{p=1}^m H_p)^2}$
 $H_{DALP} \leftarrow \frac{C_1}{-slope} + H_1$
Return H_{DALP}

Function D-DALP()

$list \leftarrow$ get_list($H_{min}=1$)
 $H_{DALP} \leftarrow$ DALP($list$)
 $list \leftarrow$ get_list($H_{min}=H_{DALP} + 1$)
 $H_{D-DALP} \leftarrow$ DALP($list$)
Return H_{D-DALP}

しない完全二部グラフ構造を持つ Boltzmann Machine のことである。

ここで、ベクトル v, h を以下のように定義する。

$$v = (v_0, v_1, \dots, v_{V-1}), \forall v_i \in 0, 1 \quad (1)$$

$$h = (h_0, h_1, \dots, h_{H-1}), \forall h_i \in 0, 1 \quad (2)$$

また、これらの結合確率 $p(v, h; \theta)$ を以下のように定義する。

$$p(v, h; \theta) = \frac{1}{Z(\theta)} \exp(-E(v, h; \theta)) \quad (3)$$

$$E(v, h; \theta) = -\sum_i b_i v_i - \sum_j c_j h_j - \sum_i \sum_j v_i W_{ij} h_j \quad (4)$$

$$Z(\theta) = \sum_v \sum_h \exp(-E(v, h; \theta)) \quad (5)$$

ここで b_i, c_j, W_{ij} はそれぞれ可視素子 v_i のバイアス、隠れ素子 h_i のバイアス、可視素子 v_i と隠れ素子 h_i の結合荷重である。 $E(v, h; \theta)$ はエネルギー関数でありこれを最小化するようにモデルのパラメータ $\theta(b_i, c_j, W_{ij})$ が更新される。今回は更新量の決定に Contrastive Divergence (CD) 法を使用した[3]。 $Z(\theta)$ は正規化定数である。

2.2 RBM の学習率自動決定法 [4]

Hinton は文献[4]において、学習率を式 (6) を用いて決定することを推奨している。

$$lr = \frac{\sum_v \sum_h^H \text{abs}(W_{vh})}{\sum_v \sum_h^H \text{abs}(\Delta W_{vh}) * R_{div}} \quad (6)$$

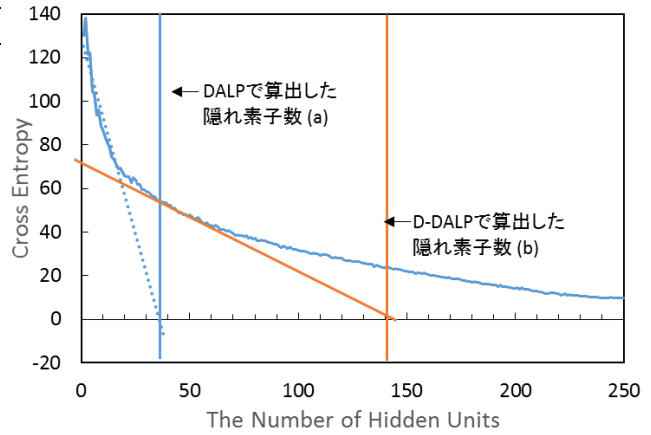


図1 DALP 法と D-DALP 法

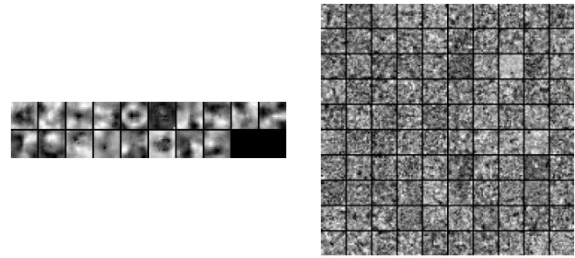


図2 隠れ素子数による獲得表現の差
素子数: 18(左) 250(右)

式 (6) は RBM の荷重結合の絶対値とその更新量を見て、荷重結合が更新量の R_{div} 倍程度になるように調整することを意味する。Hinton は文献にて二値を扱う RBM においては R_{div} を 1000 程度の値とすることを推奨している。本論文では、データセットによって R_{div} を調整する。

2.3 RBM の隠れ素子数自動決定法 (DALP 法) [1]

DALP 法は RBM の学習特性分析を行った結果得られた傾向を基にデータセットの学習に必要な隠れ素子数を自動決定する手法である[1]。疑似アルゴリズムを 1 に示す。 H は観測する隠れ素子数のリスト (2 要素以上) である。

DALP 法はまず、隠れ素子数に対してクロスエントロピーの低下が認めらる m 種類の隠れ素子数 H_p ($1 \leq p \leq m, H_p > H_l$) を等間隔で選ぶ。次に、 H_p を有した学習済みモデルのクロスエントロピーを C_p として保存する。

得られた H_p, C_p の組み合わせから最小二乗法で隠れ素子数とクロスエントロピーの関係性を線形近似し、最終的にクロスエントロピーが 0 に到達しうる隠れ素子数 H_{DALP} を算出する (図 1a)。

2.4 RBM の隠れ素子数による獲得表現の差

図 2 は異なる隠れ素子数を持つ RBM が獲得した特徴表現を可視化したものである。

隠れ素子数が 18 の RBM は少ない隠れ素子数で

表 1 実験条件

パラメータ	設定値		
	MNIST	CalTech 101 Silhouettes	Logistic Regression
学習率 lr	$R_{div} = 1000$	$R_{div} = 2500$	0.1
学習回数	50	500	1000
Batch Size	50	100	Data Set Batch Size
H_1	2	1	N/A
H_m	32	11	N/A

表 2 実験結果 (MNIST)

素子数	階層毎の識別エラー率			
	1 階層	2 階層	3 階層	4 階層
H_{DALP}	41	4.15%	3.85%	4.09%
$0.5H_{D-DALP}$	615	1.74%	1.64%	1.82%
H_{D-DALP}	1230	1.69%	1.66%	1.54%
$2H_{D-DALP}$	2460	1.73%	1.75%	1.87%

表 3 実験結果 (CalTech 101 Silhouettes)

素子数	階層毎の識別エラー率			
	1 階層	2 階層	3 階層	4 階層
H_{DALP}	53	43.78%	48.56%	65.65%
$0.5H_{D-DALP}$	200	30.52%	29.35%	31.30%
H_{D-DALP}	501	29.30%	29.35%	29.04%
$2H_{D-DALP}$	1002	28.87%	28.52%	27.48%

多数のデータを表現できるように抽象度の高い特徴を学習し、隠れ素子数が 250 の RBM は隠れ素子数の増加により局所的、つまり抽象度の低い特徴を獲得した事が示唆されている。

3. Double DALP 法

本論文にて提案する Double DALP (D-DALP) は DALP 法にて算出された隠れ素子数で学習を終えたネットワークに対し、再度 DALP 法を適用し隠れ素子数を算出する手法である (図 1b)。

D-DALP 法の疑似アルゴリズムを Algorithm 1 に示す。このアルゴリズム内の $H_{DALP} + 1$ とは DALP 法にて算出した隠れ素子数に 1 足した隠れ素子数である。H は関数 `get_list()` によって下記のように決定される。

$$\mathbf{H} = \text{get_list}(H_{min}) = (H_1, H_2, \dots, H_p, \dots, H_m), \forall H_p \geq H_{min} \quad (7)$$

図 3 は DALP 法を採用した学習済み RBM における隠れ素子数とクロスエントロピーの関係を表したグラフである。隠れ素子数の増加によってクロスエントロピーが低下していることから、本論文で扱ったデータセットでは、DALP 法にて算出された隠れ素子数のみでは特徴を獲得しきれていないことを示唆している。

D-DALP 法は DALP 法のみでは素子数の不足で学習しきれなかった特徴も学習できる素子数を算出する。また、D-DALP 法にて算出された隠れ素子数は深

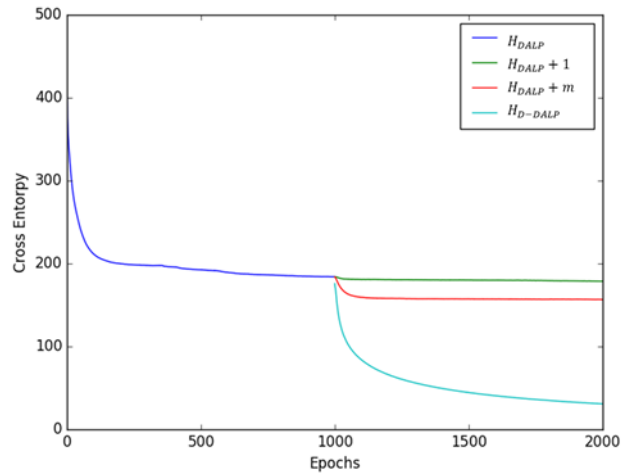


図 3 D-DALP 法のクロスエントロピーの変化

層学習の事前学習として RBM を使用するにあたって適切なパラメータであると考えられる。

4. 実画像データセットを用いた D-DALP の評価

本章では 3 章で提案した D-DALP 法を用いた学習の精度を評価するために MNIST 及び CalTech 101 Silhouettes データセットを用いた実験を行った。MNIST は手書き数字を連続値で表現した 28 ピクセル四方の画像で構成されたデータセットである。本実験では学習データセットとして 50,000 件、バリデーションデータセットとして 10,000 件、テストデータセットとして 10,000 件をそれぞれ使用した。CalTech 101 Silhouettes は CalTech 101 を二値で表現した 28 ピクセルの画像で構成されたデータセットである。今回の実験では学習データセットとして 6,324 件、テストデータセットとして 2,307 件をそれぞれ使用した。

4.1 学習特徴の識別問題における影響

DALP 法と D-DALP 法を用いた RBM がそれぞれ学習した抽象度の違う特徴が識別問題の精度にどのような影響を及ぼしたのかを分析した。今回の実験では事前学習として RBM の学習を行い、それを 1 から 4 層重ねたものにロジスティック回帰モデルを結合しファインチューニングを行った。実験条件を表 1 に示す。

MNIST の実験結果を表 2 に、CalTech 101 Silhouettes の実験結果を表 3 に示す。0.5 H_{D-DALP} と 2 H_{D-DALP} のネットワークはそれぞれ D-DALP 法で算出された隠れ素子数 H_{D-DALP} の半分と倍の素子数を使用した。示されているデータは各データセットに対して 2 回ずつ精度を観測した際の最低値である。

各 RBM のうち, MNIST の識別問題においては D-DALP 法にて導出された隠れ素子数が 2 階層を用いた場合を除き最高精度を示した。しかし, CalTech 101 Silhouettes では $2H_{D-DALP}$ が最高の識別精度を示した。

2 階層以上を採用したネットワークにおいて 2 階層目の入力素子数は 1 階層目の隠れ素子数であるため, クロスエントロピーが大幅に上昇する傾向が見られた。そのため, 抽象度がそれなりに低い特徴を学習しており, 隠れ素子数もあまり多くない $0.5H_{D-DALP}$ が二階層のネットワークにおいて高い識別精度を示したのだと思われる。

また, DALP 法にて算出した隠れ素子数を持つ RBM において階層毎の精度の差が CalTech 101 Silhouettes の 3 と 4 階層を除いてあまり示されなかった。深層学習が識別問題において極めて高い精度を示す理由として特徴抽出の階層化が挙げられる。しかし, 隠れ素子数が少ない RBM は初期の階層ですでに抽象度の高い特徴を抽出してしまうため階層化による恩恵を受け辛いためだと考えられる。以上のことから深層学習において事前学習の初期段階における学習特徴は抽象度が低いほうが精度向上に貢献すると考えられる。

実験結果から, D-DALP 法は必ずしも深層学習に適した隠れ素子を的確に算出するわけではないが DALP 法にて算出された隠れ素子数に比べ, 比較的高い精度を示す隠れ素子数を算出可能であることが示唆された。

5. おわりに

本論文では DALP 法の拡張アルゴリズムである D-DALP 法を提案した。

D-DALP 法は, DALP 法では隠れ素子数が不足しているため学習できない抽象度の低い特徴も学習可能な隠れ素子数を算出する。そのため, 識別問題の深層学習における事前学習として RBM を使用する際に比較的高精度を示す隠れ素子数を自動で決定できる。

また, DALP 法では大規模なネットワークを構築するのに IL-RBM を用いて未学習データが検出された場合のみにネットワークの拡張を行うのに対し, D-DALP 法は通常の RBM で深層学習に比較的適している隠れ素子数が算出可能であることが実験結果から示唆された。

参考文献

[1] Masahiko Osawa and Masafumi Hagiwara, Analyses of Learning Characteristics of RBMs and an Automatic Method for Deciding the Number of Hidden Units, *International Symposium on Advanced Intelligent*

Systems, F1a-5, Mokpo, Korea, 2015.

- [2] 大澤正彦, 萩原将文: RBM における未学習データ検出法の提案と追加学習への応用, 信学技報, vol. 114, no. 515, NC2014-118, pp. 283-288, 2015 年 3 月.
- [3] Carreira-Perpinan, Miguel A., and Geoffrey E. Hinton.: On Contrastive Divergence Learning. *AISTATS*. Vol. 10. 2005.
- [4] Hinton, G. E.: A Practical Guide to Training Restricted Boltzmann Machines., *Technical Report UTML TR 2010-003*, University of Toronto, pp. 599-619 (2010)