# 深層学習における正則化へのドロップアウトデザインの適用

熊澤 努 株式会社 SRA 地嵜 頌子 大阪工業大学

中川 智之 東京理科大学

kumazawa@sra.co.jp

shoko.chisaki@oit.ac.jp

1

t\_nakagawa@rs.tus.ac.jp

室井 浩明 ヱヂリウム株式会社 muroi@edirium.co.jp 渡邉 卓也 ヱヂリウム株式会社 sodium@edirium.co.jp

#### 要旨

深層ニューラルネットワークの汎化性能を向上させる 正則化法に、学習時に一部のノードだけを活性化するド ロップアウト法が知られている. ドロップアウト法が活 性化するノードを無作為に選択する技法なのに対して, ノードと重みを1エポックの訓練で均一に選択する、ド ロップアウトデザインと呼ばれる組合せ構造が提案され ている.しかしながら、正則化の効果については不明点 が多く、実験を通じて解明されることが望まれてきた. 本論文では、ドロップアウトデザインの性質を明らかす るために、ドロップアウトデザインを用いた正則化法を 多層パーセプトロンに適用した. 4種類のネットワーク 構成について,画像分類問題の精度と損失を評価する実 験を行い、すべてのネットワークについてドロップアウ ト法と同程度の汎化性能を示すことを確認した.一方で, ドロップアウトデザインに固有の特徴的な結果は観察さ れなかった.以上の成果は、ドロップアウト法とドロッ プアウトデザインをより深く理解するための手掛かりを 今後の研究に提供することが期待される.

# 1. はじめに

近年,深層学習に関連した研究開発が非常に活発に行われており,画像認識や自然言語処理をはじめとして様々な分野への応用が図られている.深層学習は訓練データに対する最適化を実行するため,訓練データに対しては高い評価性能を示す一方,未知のデータに対する汎化性

能は向上しなくなる問題が指摘されている.この過学習の問題を緩和する方法が正則化である.

ドロップアウト法 [1, 2] は最も有力な正則化法の一つ として広く知られている.ドロップアウト法は、ネット ワークの訓練の間に無作為にノードを不活性にすること で、一部の重みパラメータの学習のみを進める方法であ る.不活性にするノードはミニバッチごとに切り替える. 訓練を終了した後の、未知データに対する出力値の推定 の際には、全てのノードを活性化させる.ノードの不活 性化は、部分ネットワークで重みパラメータの値を学習 することと解釈できる.そのため、ドロップアウト法は 複数のモデル間でパラメータを共有するアンサンブル学 習の一種としても理解されている.ドロップアウト法を 使用する際には、ノードを不活性にする確率をハイパー パラメータで与える必要がある.

地寄らは、ドロップアウト法に代わる正則化技法を想 定した組合せ構造であるドロップアウトデザイン [3] を 提案した.ドロップアウトデザインは、ミニバッチごと の訓練で使用する部分ネットワーク構成の集合である. この部分ネットワークの構成をスーパーブロックという. ドロップアウトデザインは、1エポックあたりのノード の活性化回数と重みパラメータの更新回数が所与の値で 釣合うように、スーパーブロックを構成する.したがっ て、訓練時に特定のノードだけが活性化されて学習が進 む現象を緩和することが期待できる.文献 [3] では、特 定のドロップアウトデザインの構成が示されており、 計算機を用いたデザインの構成が可能となっている.そ の一方で、正則化の効果をはじめ、ドロップアウトデザ インの性質については明らかになっていない点が多い. 特に,ドロップアウトデザインを機械学習に実際に適用 してその性質を調べる研究はこれまでなされておらず, 評価実験に基づいた研究の成果が望まれる状況にある.

本論文の目標は、ドロップアウトデザインを用いた正 則化を深層学習に対して適用して、デザインの性質を実 験的に明らかにすることにある.実験では、画像分類問 題をとりあげた.ニューラルネットワークは、多層パー セプトロン (MLP)を対象とした.全結合層(隠れ層) の層数とノード数が異なる計4種類のネットワークに対 して、正則化を実行しない場合、従来のドロップアウト 法を適用した場合、ドロップアウトデザインを適用した 場合についてそれぞれ汎化性能を評価した.その結果、 次の二点を確認した.

- すべてのネットワークについて、ドロップアウト法 とドロップアウトデザインによる正則化法は同程度 の精度と損失を示した。
- ドロップアウト法にはない特徴的な結果は、ドロップアウトデザインを用いた結果には見られなかった.

以上の結果は、ノードを無作為に選ぶドロップアウト 法と、ブロックによって均一にノードを活性化させるド ロップアウトデザインには、性能上大きな差がないこと を示唆している. このことは、組合せ構造であるドロッ プアウトデザインと無作為性を持つドロップアウト法の 関連性を示しており、今後の研究において、組合せ論を 使ったことによりドロップアウト法を分析するアプロー チが期待できる.本論文のもう一つの貢献は、ドロップ アウトデザインをニューラルネットワークに適用する具 体的な方法を示したことである. ドロップアウトデザイ ンの適用に際しては、デザインの構造とニューラルネッ トワークの構成との対応づけ、複数エポックの訓練にお けるスーパーブロックの割り当てを決定する必要がある. 本論文では、多層パーセプトロンの全結合層に対して正 則化を行う場合のドロップアウトデザインの適用法を示 し,実験により汎化性能の違いを調べた.

本論文の構成は次の通りである.2節で研究の背景と なるドロップアウト法とドロップアウトデザインについ て説明する.3節では、ドロップアウトデザインの深層 学習への適用法を説明する.4節で、評価実験の結果を 報告する.5節で関連する研究を論じ、6節で結論と今 後の課題を述べる.

# 2. 背景

ここでは、本論文が着目するドロップアウト法とド ロップアウトデザインの概要を説明する.

#### 2.1. ドロップアウト法

ドロップアウト法 [1, 2] は、ニューラルネットワーク のノードを無作為に不活性にして訓練を行うことで、汎 化性能を高める技術である. ノードを不活性にする確率 であるドロップアウト率pはハイパーパラメータとして 与えられる、訓練時には、層ドロップアウトマスクと呼 ばれるベクトル m で層の出力をマスキングすることで ノードの活性化と不活性化を行う [4]. ただし、mの各 成分 $m_i$ はベルヌーイ分布に従って確率pで0を、確率 1-pで1をとる. すなわち、 $m_i \sim Bernoulli(1-p)$ と する. ここで、mの各成分はノードを表し、1であれば 活性化することを、0であれば不活性化することを意味 する. ドロップアウト法を適用する層への入力 x と出力 y との間には以下の関係が成り立つ [4].

$$\mathbf{y} = f(\mathbf{W}\mathbf{x} + \mathbf{b}) \circ \mathbf{m} \tag{1}$$

ここで, W は層の重み, b はバイアス, *f* は活性化関数, ∘ は成分ごとの積を表す.未知データに対する推定時には全てのノードを活性化し,以下のように平均化を行う.

$$\mathbf{y} = (1-p)f(\mathbf{W}\mathbf{x} + \mathbf{b}) \tag{2}$$

ドロップアウト法の異なる定式化として、訓練時には (1) 式の右辺に 1/(1-p) を乗じた値を出力する一方、推定 時には (2) 式の (1-p) の乗算を実行しない方式があり、 どちらの定式化を用いても、推定時の出力が等しいこと が知られている [4]. 簡単のため、本論文では両者を区 別せず、(1) 式と (2) 式を用いる.

ドロップアウト法で訓練に使われるネットワークは, 元のネットワークの部分ネットワークである.したがっ て,ドロップアウト法を,重みを共有する複数のネット ワークによるアンサンブル学習の一種とみなすことがで きる [1].

#### 2.2. ドロップアウトデザイン

ドロップアウトデザインは,ニューラルネットワーク の重みを更新する回数を均一にすることを目的とした組 合せ構造の一種である.ドロップアウトデザインは複数 の点集合から構成される.ネットワークの各層のノード の集合を異なる点集合とみなし,層を横断した部分集合 をとることによって活性化するノードを選択するモデル を与える.V<sub>1</sub>,V<sub>2</sub>,...,V<sub>n</sub>を互いに異なる点集合とし,そ れぞれの部分集合の集合 (スーパーブロックと呼ぶ)を 以下のように定める.

 $\mathcal{B} = \{ \{ C_1 | C_2 | \cdots | C_n \} \mid C_i \subset V_i, \ C_i \neq \emptyset, \ 1 \le i \le n \}.$ 

各部分集合 $C_i$ をサブブロックと呼ぶ. 任意の $0 \le i \le n - s$ について,連続したs 個の点集合 $V_{i+1}, V_{i+2}, \ldots, V_{i+s}$ それぞれから選出した任意の $d_1, d_2, \ldots, d_s$  個の点を同時に含むスーパーブロックが $\mathcal{B}$ の中に $\lambda_i$  個存在するとき, $(V_1, V_2, \ldots, V_n; \mathcal{B})$ を $(d_1, d_2, \ldots, d_s)$ 型-ドロップアウトデザインと呼ぶ [3].  $\lambda_i$  ( $0 \le i \le n - s$ )はドロップアウトデザインの会合数と呼ばれる. 会合数を一定にすることで,ノードの活性化回数及びエッジの使用回数の均一化を実現する. 2 層のネットワークでの例を以下に示す.

**例 1.** 点集合を  $V_1 = \{0, 1, 2, 3\}, V_2 = \{0, 1, 2, 3, 4, 5\}$ とする. このとき,  $(V_1, V_2; \mathcal{B})$  は会合数  $\lambda_1 = 1$  をもつ (2, 1) 型-ドロップアウトデザインをなす. ただし,

$$\begin{split} \mathcal{B} &= \{ \{0,1 \mid \mathbf{0},\mathbf{1},\mathbf{2}\}, \quad \{0,1 \mid \mathbf{3},\mathbf{4},\mathbf{5}\}, \quad \{0,2 \mid \mathbf{1},\mathbf{2},\mathbf{5}\}, \\ &\{0,2 \mid \mathbf{0},\mathbf{3},\mathbf{4}\}, \quad \{0,3 \mid \mathbf{1},\mathbf{2},\mathbf{4}\}, \quad \{0,3 \mid \mathbf{0},\mathbf{3},\mathbf{5}\}, \\ &\{1,2 \mid \mathbf{0},\mathbf{2},\mathbf{3}\}, \quad \{1,2 \mid \mathbf{1},\mathbf{4},\mathbf{5}\}, \quad \{1,3 \mid \mathbf{0},\mathbf{4},\mathbf{5}\}, \\ &\{1,3 \mid \mathbf{1},\mathbf{2},\mathbf{3}\}, \quad \{2,3 \mid \mathbf{0},\mathbf{1},\mathbf{3}\}, \quad \{2,3 \mid \mathbf{2},\mathbf{4},\mathbf{5}\} \}. \end{split}$$

このデザインでは,例えば, $V_1$ の2個の点0,1と $V_2$ の 点**0**を同時に含むスーパーブロックは1個だけである.

先行研究において, 直交配列を用いた構成法や, アフィ ン空間や射影空間の有限幾何を用いた構成法が提案され ている.本論文の実験に使用したアフィン空間による構 成法は, (2,1)型かつ (1,2)型-ドロップアウトデザイン を構成することができる (Theorem 5.13 [3]). ここで, デザインから構成されるモデルは,素数ベキ q,整数 t,  $d \ge 3$  について, 層数は  $q^{d-t}$ , 各層のノード数は  $q^t$  であ り, 会合数は各 i に対して  $\lambda_i = (q^{d-2} - q^{d-t-1})/(q-1)$ をとる.また, 各層のノードを活性化する割合は 1/q で ある.したがって, 不活性にするノードの割合, すなわ ちドロップアウト率は 1 - 1/q となる.

### 3. ドロップアウトデザインの深層学習への適用

本節では、ドロップアウトデザインのニューラルネットワークへの適用方法について述べる.

本論文では、訓練に使用するドロップアウトデザイン をネットワークごとに固定する.ドロップアウトデザイ ンは、1エポックの学習に必要な部分ネットワークの集 合を定めたものと考える.2.2節で述べたように、ドロッ プアウトデザインの点集合は、ノードの集合と1対1に 対応させる.スーパーブロックは、1つのミニバッチで学 習を進める部分ネットワークであり、サブブロックは各 層において活性化するノードの集合とする.(1)式にお いて、サブブロックに含まれる点に対応づけられたノー ドを1、含まれない点に対応づけられたノードを0とす るように層ドロップアウトマスク mを構成することで、 訓練時のノードの活性化と不活性化を実現する.

ドロップアウトデザインの各スーパーブロックを、1エ ポックにつき1回、重みの更新に使用する.ただし、複数 エポックの訓練におけるスーパーブロックの使用法につ いては、ドロップアウトデザインの定義には定められて いない.本論文では、第1エポックで使用するスーパー ブロックの順序を任意に一つ固定するものとし、複数エ ポックの訓練について、次の二通りの使用法を考える.

- 第1エポックで定めたスーパーブロックの使用順序 を巡回的に保つ.第2エポック以降では、各エポッ クの開始時にスーパーブロックを1だけシフト移動 して巡回したデザインを構成することで、スーパー ブロックの使用順序を定める.以降では、この使用 方法をブロックシフトと呼ぶことにする.ブロック シフトにより構成されるドロップアウトデザインは 互いに同値である点に注意する.
- 第1エポックで定めたスーパーブロックの使用順序 を第2エポック以降も引き続き使用する.

**例 2.** 例 1 のドロップアウトデザイン *B* を再び考える. 表1にブロックシフトの適用例を示す.まず,*B* のスー パーブロックの集合を順列と読み替えることで,第1エ ポックでのスーパーブロックの使用順序を表 1a のよう に定める.このデザインにブロックシフトを適用すると, 第2エポックでのスーパーブロックの使用順序が表 1b のように定まる.ブロックシフトを適用しない場合には, 表 1a の順序を第2エポック以降も使用する.

表 1: ブロックシフトの例

(a)	第1エ:	ポック		<b>(b)</b> 第2エポック					
順序	1層	2 層	]	順序	1層	2 層			
1	0, 1	0, 1, 2		1	2, 3	2, 4, 5			
2	0, 1	<b>3</b> , <b>4</b> , <b>5</b>		2	0,1	<b>0</b> , <b>1</b> , <b>2</b>			
3	0, 2	<b>1</b> , <b>2</b> , <b>5</b>		3	0,1	$\mathbf{3, 4, 5}$			
4	0, 2	0, 3, 4		4	0,2	${\bf 1,2,5}$			
5	0, 3	<b>1</b> , <b>2</b> , <b>4</b>		5	0,2	0, 3, 4			
6	0, 3	<b>0</b> , <b>3</b> , <b>5</b>		6	0,3	${\bf 1,2,4}$			
7	1, 2	0, 2, 3		7	0,3	0, 3, 5			
8	1, 2	$\mathbf{1, 4, 5}$		8	1,2	0, 2, 3			
9	1, 3	$\mathbf{0, 4, 5}$		9	1,2	${\bf 1, 4, 5}$			
10	1, 3	<b>1</b> , <b>2</b> , <b>3</b>		10	1,3	$\mathbf{0, 4, 5}$			
11	2, 3	<b>0</b> , <b>1</b> , <b>3</b>		11	1,3	1, 2, 3			
12	2,3	${f 2}, {f 4}, {f 5}$		12	2, 3	0, 1, 3			

文献 [3] ではドロップアウトデザインを適用するネットワーク構成について特に仮定は設けられていない.本 論文では,MLPを対象として,これらのネットワーク の全結合層にドロップアウトデザインを適用する.その ためには,各層の出力ベクトルの各成分とデザインの各 点とを一対一に対応づければよい.

# 4. 評価実験

本節では、計算機による評価実験の結果を報告する. 本実験の目的はドロップアウト法に対する優位性を評価 することではなく、ドロップアウトデザインによる正則 化とドロップアウト法による正則化の効果に違いがある か比較検討することで、ドロップアウトデザインの性質 を調べることである.

#### 4.1. 実験設定

本実験では,深層ニューラルネットワークの全結合層 に対して,訓練時の性能と,テストデータを用いた時の 汎化性能の評価を行った.性能評価は,ドロップアウト デザインを用いた正則化を施した場合,ドロップアウト 法を用いた正則化を施した場合,正則化を行わない場合 を対象とした.

本実験で扱う問題は画像分類問題とした. データセット には CIFAR-10 [5] を使用した. CIFAR-10 は, 50,000 個

4

の訓練データと 10.000 個のテストデータからなる 10 ク ラス画像分類用データセットである. 各データは 32×32 ピクセルのカラー画像である.3節で述べたように、実験 を行うニューラルネットワーク構成は MLP とした.ド ロップアウトデザインとドロップアウト法はネットワー クの全結合層だけに適用した.使用するドロップアウト デザインは、4種類の(2,1)かつ(1,2)型-ドロップアウ トデザインとした. 全結合層の構成は、実験で使用する ドロップアウトデザインに基づいて定めた.表2に、実 験で扱うドロップアウトデザインと、対応する MLP の 全結合層の構成を示す.4種類の各ネットワークに対し てドロップアウト法を適用する場合には、各層のドロッ プアウト率を表2にある値に設定した.この値は、ド ロップアウトデザインによって定まるノードの不活性率 と同一の値である. ドロップアウトデザインを適用する 際には、ブロックシフトを実行する場合と、ブロックシ フトを行わず、どのエポックでも同一順序でスーパーブ ロックを適用する場合の二通りを評価した. スーパーブ ロックの使用順序の性能への影響を調べるためである.

各ネットワークの損失関数には交差エントロピーを, パラメータの更新には確率的勾配降下法(SGD)を使用 した.また,入力層と全結合層の活性化関数は ReLU 関 数とし,出力層の活性化関数は Softmax 関数とした.

実験では、各ネットワーク構成について、500 エポッ クの学習を行った.その後、テストデータに対する精度 (正解率)と、交差エントロピーに基づく損失を評価し た.精度と損失はテストデータの正解クラスについて求 めた.この実験をそれぞれのネットワーク構成について 10回ずつ行った結果を評価した.ドロップアウト法とド ロップアウトデザインの正則化効果だけを識別しやすく するため、データ拡張などの他の汎化性能を向上させる 技法は使用しなかった.

実験プログラムは Python 3 で実装した. ニューラル ネットワークの訓練ならびにテストの実装には, Keras ライブラリを用いた. ドロップアウト法には, Keras が 提供するドロップアウト層の実装を使用した. 一方, ド ロップアウトデザインに関しては, アフィン空間を利用 したドロップアウトデザインの構成法 (2.2 節)を実装 した. また, Keras のドロップアウト層と同じ実装方法 でデザインを用いたノードの活性化を実現するために, 従来のドロップアウト層を拡張したデザイン適用層を新 たに実装した. このデザイン適用層を用いて, 実験プロ グラムは, 構成したデザインから層ドロップアウトマス

アフィン空間			ドロップアウトデザイン				ニューラルネットワーク(全結合層)				
		flat	サブ	点	サブ	スーパー	会	層	1層の	1層の	ドロップ
次元	位数	の	ブロック	集合	ブロック	ブロック	合		ノード	活性化	アウト
d	q	次元 t	数	サイズ	サイズ	数	数	数	数	ノード数	率
8	2	7	2	128	64	508	63	2	128	64	0.5
7	3	6	3	729	243	3276	121	3	729	243	0.67
9	2	7	4	128	64	1016	126	4	128	64	0.5
10	2	8	4	256	128	2040	254	4	256	128	0.5

表 2: 実験で使用したドロップアウトデザインとネットワーク構成の対応

クを生成して,(1)式に従って訓練を実行する.テスト 時には,ドロップアウト法と同様に(2)式により出力を 計算することとした.

実験には、CPU に Intel Xeon W-2133 (3.60GHz), GPU に NVIDIA TITAN RTX を備えたマシン(Keras 2.3.1, Tensorflow 2.1.0)と、CPU に Intel Xeon W-2245 (3.90GHz), GPU に NVIDIA TITAN RTX を備えたマ シン(Keras 2.4.3, Tensorflow 2.3.0)を使用した.

#### 4.2. 実験結果

MLP の全結合層を2層,3層,4層にした場合の実験 結果をそれぞれ図1,図2,図3,図4に示す.測定し た値は,10回の実験に対する精度の平均と標準偏差,損 失の平均と標準偏差である.加えて,第51エポックから 第500エポックまでの各エポックについて,そのエポッ クを含む過去50エポックの区間に対する精度と損失の 分散を算出した.これは,訓練の進行に伴う,実験ごと の汎化性能の収束の違いを評価する指標である.各グラ フは,正則化を行わなかった結果(w/o dropout),ド ロップアウト法による正則化を行った結果(dropout), ブロックシフトを適用した場合(design(bs))と適用し ない場合(design)のドロップアウトデザインでの正則 化の結果をそれぞれ示している.グラフの横軸はエポッ ク数である.

図1から図4から、以下の三点が観察された.

 本実験では、ドロップアウトデザインを用いること で汎化性能が高まるかどうかを調べるため、精度と 損失の平均、標準偏差、分散を測定した.それらす べての項目について、ドロップアウトデザインを用 いた正則化は、従来のドロップアウト法と同程度の 汎化性能を達成した.

- 無作為にノードを不活性にする従来のドロップアウト法と異なり、ドロップアウトデザインは活性化するノードの回数に制約を設けた決定論的な組合せ構造である。そのため、ドロップアウトデザインの固有の特徴を示す結果が得られることが期待される。 実験の結果からは、精度と損失に関して、ドロップアウトデザインとその適用法に特有の性質や挙動は確認できなかった。
- ドロップアウトデザインの適用において、ブロック シフトを実行した場合としない場合を比較したと き、精度と損失の平均、標準偏差に大きな違いは確 認されなかった.この結果は、スーパーブロックの 使用順序は汎化性能に影響しないことを示している と考えられる.

ドロップアウト法との正則化の違いを評価するという 点から、まず、上で述べた結果1と2を詳しく検討する. 本実験の結果から、表2に示したすべてのネットワー ク構成について、ドロップアウト法とドロップアウトデ ザインを用いた正則化は平均的に同程度の精度と損失 であり、各曲線の収束の仕方にも大きな違いは見られな かった.ただし、精度と損失の平均値に関しては、全結 合層が4層,各層のノード数が128の場合には、両者の 挙動にわずかながら違いを確認した. 訓練データ、テス トデータ共に平均精度はドロップアウト法が上回り(図 3a, 3b), ドロップアウト法が低損失になるという傾向 が観察された(図 3c, 3d). また,同じ構成について, 精度の標準偏差の結果(図 3e, 3f)から、ドロップアウ ト法を用いた場合は、ドロップアウトデザインを用いる 場合よりもわずかに低い標準偏差を示した. このような 結果が得られた原因として、中間層数が最も大きいネッ トワークを使用していることが考えられる.しかし、各 層を256ノードとした結果(図4e,4f)では同様の傾向 は見られないことから、ドロップアウトデザインの固有 の特徴を示す結果とはいえないと思われる.ニューラル ネットワークの訓練とドロップアウト法における確率的 な挙動が影響している可能性もあり、本論文で行った実 験だけからでは結論を得ることは難しい.今後、5層以 上の深いネットワーク対する評価実験が必要である.50 エポック区間の精度分散,損失分散に関しては、全結合 層2層、3層、4層のいずれの場合も、ドロップアウト 法とドロップアウトデザインを用いた正則化と顕著な違 いは見られなかった.以上の考察から、訓練の進行に伴 う精度や損失の収束の仕方には大きな違いはないと結論 付けられる.

次に、ドロップアウトデザインの結果に注目して、ブ ロックシフトを実行した場合としない場合を比較した場 合を議論する.上の3で述べたように、すべての実験を 通じて、精度と損失の平均、標準偏差に大きな違いは確 認されなかった.なお、全結合層4層、各層128ノー ドに対する50エポック区間の精度と損失の分散(図3i, 3j)において、ブロックシフトの方が訓練の初期段階で の分散値が低くなる傾向が見られるが、その差は微小で ある.

最後に、実験結果への影響を与える要因として、過学 習の可能性と収束速度について検討する. 正則化を行わ ない場合については、全結合層3層(図2a)と4層かつ 各層 256 ノード(図 4a)のそれぞれにおいて、短いエ ポック数の訓練で精度が1.00に達して収束し、また、標 準偏差が0になった.そのため、訓練の進行と共に過学 習を引き起こしている可能性が高い. 全結合層4層,各 層 128 ノードの場合は、図 3e においても、エポック数の 増加に伴いテストデータに対する精度が低下する傾向が あり、過学習となっている恐れがある. これらのネット ワーク構成では, 正則化技法を用いない場合に最も短い 収束速度で高い汎化性能が得られた, という結果を示し ている.以上の議論から、この現象は過学習が原因であ る可能性がある. ドロップアウト法とドロップアウトデ ザインを用いた場合には,過学習の傾向は見られなかっ た. 全結合層3層と4層については、訓練時の平均精度 が収束まで達しておらず, 正則化技法を用いない場合と 比較して、収束に遅れが見られる. 訓練するエポック数 をさらに増やすと、精度や損失の結果に違いが現れる可 能性がある.

#### 5. 関連研究

ドロップアウト法は後に続く正則化法の研究に大きな 影響を与えた. 代表的な正則化法の一つにドロップコネ クト法 [6] がある. ドロップコネクト法は、学習時にノー ド間の接続を無作為に切断する方法である.一方,高速 ドロップアウト法 [7] は、ガウス近似によりドロップア ウト法と同様の正則化効果を発生させることで、学習時 の収束速度を高める方法である.より最近では、畳み込 み層の出力の周波数成分に対してドロップアウト法を適 用するスペクトラルドロップアウト法 [8]、ドロップアウ ト法を用いた訓練をベイズ推定で近似する MC ドロップ アウト法 [9, 10], 学習効果の高い部分ネットワークを進 化計算を用いて絞り込む EDropout [11], 強化学習でド ロップアウトマスクを学習する AutoDropout [12] など の多くの方法が提案されている. また, R-Drop は、ド ロップアウト法で部分ネットワークを二つ構成し、両者 が推定した分布の差異を低減するように訓練を行う正則 化法である [13]. ドロップアウト法と深く関係する主要 な技法は文献 [4] で詳しく論じられている.

組合せデザインをはじめとする組合せ構造の計算機科 学への応用には、数多くの研究がある [14]. デザインに よるバギング [15, 16] は、バギングにおいて組合せデザ インの考え方に基づきデータをサブサンプリングするこ とで、汎化性能を改善する技法である. バギングは複数 の弱分類器で学習を行うアンサンブル学習の一種である. ドロップアウト法もアンサンブル学習と考えることもで きるので、デザインによるバギングは、組合せデザイン を機械学習に適用する先駆的な試みの一つである.また、 組合せデザインと関連が深い分野に実験計画法や品質工 学(タグチメソッド)があり、ニューラルネットワーク の最適な構成を求める技法 [17, 18] をはじめ、機械学習 での活用が研究されている.

### 6. おわりに

6

深層学習における正則化に関しては、数多くの研究 がなされてきた.本論文で扱ったドロップアウトデザイ ンを用いた正則化技法は、部分ネットワークの組合せ構 造に注目した方法である.無作為にノードを不活性にす るドロップアウト法とは異なり、ドロップアウトデザイ ンは、ブロックによってノードの活性化回数を釣り合わ せるという特徴を持つ.本論文では、ドロップアウトデ ザインによる正則化の持つ性質を明らかにするために, MLP の全結合層を正則化する技術を実装した.加えて, 画像分類問題での評価実験により,全結合層が2,3,4 層の時の汎化性能を分析した.その結果,いずれの場合 についても,ドロップアウト法と同程度の汎化性能が得 られた一方で,ドロップアウトデザイン固有の特徴的な 結果は得られなかった.これらのことは,訓練時に使用 したドロップアウトデザインが持つ組合せ論に基づいた ネットワーク構造と,従来のドロップアウト法で得られ る無作為な構造との間に,関係性があることを示唆して いると考えられる.ドロップアウトデザインの分析を通 じて,組合せ構造を用いてドロップアウト法を理解する 研究が進むものと期待される.ドロップアウトデザイン のニューラルネットワークへの適用法の違いは,実験結 果に大きな影響を与えることはなかった.

今後、ドロップアウトデザインとドロップアウト法に ついてさらに深く理解するために,以下の課題に取り組 む必要がある.まず,引き続きより深いニューラルネッ トワークや、画像分類以外の問題についても実験を進め ていく必要がある.また、本論文での評価実験は(2,1) かつ (1,2)型-ドロップアウトデザインに限った結果であ り、ドロップアウトデザインの型が汎化性能に与える影 響を分析する必要がある.文献 [19] のように,ドロップ アウトデザインの構成法の研究は進展しているが、様々 な型のドロップアウトデザインを理論的に構成すること は現時点では困難である.そのため、従来のドロップ後 法と異なり、ドロップアウト率を自由に設定して訓練を 行うことができないという制約がある.そこで、計算機 によりドロップアウトデザインを構成する技術を開発す ることが必要である.計算機を利用して組合せ構造を探 索する研究 [20, 21, 22] が参考になると考えられる. 最 後に、本論文の実験では、ニューラルネットワークのド ロップアウトデザインへの適用法の有効性を確認できな かった. 今後も有効な適用法を検討する必要がある.

### 謝辞

本研究は JSPS 科研費 基盤研究 (C) 19K11866 及び, 若手研究 21K13845 の助成を受けたものです.

### 参考文献

[1] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580*, 2012.

- [2] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [3] Shoko Chisaki, Ryoh Fuji-Hara, and Nobuko Miyamoto. Combinatorial designs for deep learning. *Journal of Combinatorial Designs*, 28(9):633–657, 2020.
- [4] Alex Labach, Hojjat Salehinejad, and Shahrokh Valaee. Survey of dropout methods for deep neural networks. arXiv:1904.13310, 2019.
- [5] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.
- [6] Li Wan, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. Regularization of neural networks using DropConnect. In 30th International Conference on Machine Learning, volume 28 of Proceedings of Machine Learning Research, pages 1058–1066. PMLR, 2013.
- [7] Sida Wang and Christopher Manning. Fast dropout training. In 30th International Conference on Machine Learning, volume 28 of Proceedings of Machine Learning Research, pages 118–126. PMLR, 2013.
- [8] Salman H. Khan, Munawar Hayat, and Fatih Porikli. Regularization of deep neural networks with spectral dropout. arXiv:1711.08591, 2017.
- [9] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In 33rd International Conference on Machine Learning, volume 48 of Proceedings of Machine Learning Research, pages 1050–1059. PMLR, 2016.
- [10] Yarin Gal and Zoubin Ghahramani. A theoretically grounded application of dropout in recurrent neural networks. In Advances in Neural Information Processing Systems, volume 29, pages 1027–1035, 2016.
- [11] Hojjat Salehinejad and Shahrokh Valaee. EDropout: Energy-based dropout and pruning of deep neural networks. arXiv:2006.04270, 2020.
- [12] Hieu Pham and Quoc Le. AutoDropout: Learning dropout patterns to regularize deep networks. Proceedings of the AAAI Conference on Artificial Intelligence, 35(11):9351–9359, 2021.
- [13] Xiaobo Liang, Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Min Zhang, and Tie-Yan Liu. R-drop: Regularized dropout for neural networks. In *Thirty-fifth Conference on Neural Information Pro*cessing Systems (NeurIPS 2021), 2021.

- [14] Charles J. Colbourn and Paul C. van Oorschot. Applications of combinatorial designs in computer science. ACM Comput. Surv., 21(2):223–250, 1989.
- [15] Periklis Papakonstantinou, Jia Xu, and Zhu Cao. Bagging by design (on the suboptimality of bagging). AAAI Conference on Artificial Intelligence, 28(1), 2014.
- [16] Mahdi Hamdani, Patrick Doetsch, and Hermann Ney. Bagging by design for continuous handwriting recognition using multi-objective particle swarm optimization. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR), pages 256– 260, 2015.
- [17] F. Sánchez Lasheras, J. A. Vilán Vilán, P. J. García Nieto, and J. J. del Coz Díaz. The use of design of experiments to improve a neural network model in order to predict the thickness of the chromium layer in a hard chromium plating process. *Mathematical and Computer Modelling*, 52(7–8):1169–1176, 2010.
- [18] Ahmad M. Karim, Mehmet S. Güzel, Mehmet R. Tolun, Hilal Kaya, and Fatih V. Çelebi. A new generalized deep learning framework combining sparse autoencoder and Taguchi method for novel data classification and processing. *Mathematical Problems in Engineering*, 2018:3145947, 2018.
- [19] Shoko Chisaki, Ryoh Fuji-Hara, and Nobuko Miyamoto. A construction for circulant type dropout designs. *Designs, Codes and Cryptography*, 89(8):1839–1852, 2021.
- [20] 松中春樹; 丹生智也; 番原睦則; 田村直之. SAT 符号化を 用いた釣合い型不完備ブロック計画の構成. 人工知能学 会論文誌, 27(2):10–15, 2012.
- [21] 青柳卓; 宮本暢子; 篠原聡. Optical orthogonal codes の 探索 II. **明星大学研究紀要(情報学部)**, 22:15–22, 2014.
- [22] B. N. Mandal, Rajender Parsad, and Sukanta Dash. Construction of a-optimal balanced treatment incomplete block designs: An algorithmic approach. Communications in Statistics - Simulation and Computation, 49(6):1653–1664, 2020.



図 2: 全結合層 3 層, 1 層当たり 729 ノード,ドロップアウト率 0.67 の実験結果



図 4: 全結合層 4 層, 1 層当たり 256 ノード, ドロップアウト率 0.5 の実験結果