

Deep Learning

深層學習

人工知能学会 監修

神鳶敏弘 編

麻生英樹・安田宗樹・前田新一・岡野原大輔
岡谷貴之・久保陽太郎・ボレガラ ダヌシカ 共著

近代科学社

確定的

階層型ニューラルネット

hierarchical neural network [HNN]

深層化

再帰的部分構造

深層(階層型)ニューラルネット

deep (hierarchical) neural network [DNN]

深層化

畳み込み構造

畳み込みニューラルネット

convolutional neural network [CNN]

畳み込み入力

回帰結合ニューラルネット

recurrent neural network [RNN]

回帰結合入力

自己符号化器

autoencoder [AE]

深層化

深層自己符号化器

deep autoencoder [DAE]

深層化

積層自己符号化器
stacked autoencoder

雑音除去自己符号化器
denoising autoencoder

積層自己符号化器

深層自己符号化器

深層化

確率的

制限ボルツマンマシン

restricted Boltzmann machine [RBM]

深層化(有向)

深層化(無向)

深層化(無向)

深層信念ネットワーク

deep belief network [DBN]

深層化

深層ボルツマンマシン

deep Boltzmann machine [DBM]

深層化

深層学習

深層学習手法の全体像

Deep Learning

深層學習

人工知能学会 監修

神鳥敏弘 編

麻生英樹・安田宗樹・前田新一・岡野原大輔

岡谷貴之・久保陽太郎・ボレガラ ダヌシカ 共著

近代科学社

◆ 読者の皆さまへ ◆

平素より、小社の出版物をご愛読くださいまして、まことに有り難うございます。

(株)近代科学社は1959年の創立以来、微力ながら出版の立場から科学・工学の発展に寄与すべく尽力してきております。それも、ひとえに皆さまの温かいご支援があつてのものと存じ、ここに衷心より御礼申し上げます。

なお、小社では、全出版物に対してHCD(人間中心設計)のコンセプトに基づき、そのユーザビリティを追求しております。本書を通じまして何かお気づきの事柄がございましたら、ぜひ以下の「お問合せ先」までご一報くださいますよう、お願ひいたします。

お問合せ先：reader@kindaikagaku.co.jp

なお、本書の制作には、以下が各プロセスに関与いたしました：

- ・企画：小山 透
- ・編集：高山哲司
- ・組版：藤原印刷（LaTeX）
- ・印刷：藤原印刷
- ・製本：藤原印刷（PUR）
- ・資材管理：藤原印刷
- ・カバー・表紙デザイン：川崎デザイン
- ・広報宣伝・営業：富高琢磨、山口幸治

●本書に記載されている会社名・製品名等は、一般に各社の登録商標または商標です。本文中の©、®、™等の表示は省略しています。

- ・本書の複製権・翻訳権・譲渡権は株式会社近代科学社が保有します。
- ・[JCOPY]〈(社)出版者著作権管理機構 委託出版物〉
本書の無断複写は著作権法上での例外を除き禁じられています。
複写される場合は、そのつど事前に(社)出版者著作権管理機構
(電話 03-3513-6969, FAX 03-3513-6979, e-mail: info@jcopy.or.jp) の
許諾を得てください。

序 文

深層学習 (deep learning) は、ここ数年、機械学習（あるいは人工知能）の分野で大きな注目を集めている技術である。「深層」というのは、学習を行うニューラルネットワーク（ないしはそれに相当するもの）において層が深い、すなわち、何段にも層が積み重なっているということであり、深層学習とは多段の層を持つ機構を用いた学習のことである。深層ニューラルネットワークを構築することは、長い間、研究領域におけるある種の「夢」であった。というものも、人間の脳は、多段に積み重なった構造をしているにもかかわらず、それと同じような構造を模擬するだけでは、肝心の学習の能力がうまく発揮されないということがあったためである。

そして、この「深層」であるということは、機械学習を適用する際の「問題の表現」において極めて重要である。表現というのは、与えられたデータないしは外界のどこに注目し特徴量として表すかということであるが、これまでそれを決めるのは人間の能力に頼るほかなかった。機械が特徴量を自動的に抽出できる、すなわち表現を学習することができれば、それは画期的なことである。深層学習は、表現学習を実現する一つの有望な方法であり、人工知能の分野全体にとっての潜在的な意義も大きい。

本書は、人工知能学会の『人工知能』誌に 2013 年 5 月号から 2014 年 7 月号の間に計 7 回にわたって連載した連載解説「Deep Learning（深層学習）」に加筆と再編を行ったものである。本書でも各章ごとに完結しており、個別に読み進めることができる。深層学習に関わる研究領域の第一線で活躍している研究者が執筆している。もともと人工知能学会の会員を対象にしたものであるが、各章ともにわかりやすく書かれており、情報系の大学院生レベルであれば、理解することができる内容であろう。また、詳しい数式は追わず、その意味するところを把握するだけであれば、技術系の幅広い読者にとって十分に役立つで

あろう。そういう方には、まず全体像を把握した後、必要な箇所を、参考文献とあわせ深く理解していくことをお勧めする。

各章は以下のように構成されている。第Ⅰ部は第1章から第4章まで構成され、基礎的な内容および実装面における内容である。

第1章は麻生英樹による深層学習全体の位置づけや歴史的な経緯であり、深層学習を俯瞰的に理解するのに適している。なぜ深層か、なぜ表現かという基本的な問い合わせ始まり、幅広い視野からややもすれば過度な期待とともに語られる深層学習を客観的に解説している。冷静さと期待感を同時に含んだ、極めて正しい論が展開されており、深層学習の俯瞰的な解説として秀逸である。

第2章は安田宗樹による解説であり、深層学習における一つの大きな理論的モデルの柱であるボルツマンマシンに関して述べられている。ボルツマンマシンのモデルや学習といった基礎からはじまって、深層学習でよく用いられる制限ボルツマンマシンと話が進み、深層ボルツマンマシンが導入される。自己符号化器とあわせて事前学習の意味についても議論されている。

第3章では、前田新一により、制限ボルツマンマシンの学習に用いられるコントラスティブ・ダイバージェンス法（CD法）について解説されている。CD法の根本となる考え方、事前学習の意味や位置づけからはじまり、深層学習におけるCD法について詳しく述べられている。

第4章は、大規模な実装という観点からの岡野原大輔による解説であり、世界各国で開発競争が行われている核心部分について触れられている。GPUの活用や並列化・分散化といったテーマについて、最新のアプローチが紹介されている。実装時のチェックポイントや超パラメータの最適化といった話題も価値が高い情報であろう。

次に、第Ⅱ部では、深層学習の応用に焦点をあて、画像、音声、および自然言語と、それぞれ具体的な分野について解説されている。

第5章では、岡谷貴之による画像認識の解説であり、深層学習でも最もスポットライトを浴びる研究テーマである。画像認識で高い精度を出している畳み込みニューラルネットワークのモデルと具体例について述べた後、深層学習が注目されている要因である、教師なし学習による画像特徴の抽出について説明されている。両者の位置づけについての示唆に富んだ解説が展開される。

第6章は、画像認識と並んで、深層学習のハイライトの一つである音声認識について、久保陽太郎が解説する。画像認識との大きな違いは時系列の扱いであり、時系列を扱うためのモデルが導入され、それを深層にするための方法や事前学習について述べられている。音響モデルと同時に言語モデルについても、最新の手法も含め、解説されている。

第7章は、ボレガラ・ダヌシカによる自然言語処理分野における解説である。自然言語処理の分野は、深層学習にとっては手強い分野である。単語や文をいかに表現するかにさまざまな工夫が行われており、その中でも基本的なニューラル言語モデル、単語の分散表現などからはじまり、近年注目を集めているword2vecについての解説も含まれている。

どの章を読んでも印象的のが、深層学習はまだはじまったばかりのものであり、理論的にも不十分な点、解明されていない点が多いということである。これは裏を返せば、研究の余地や発展の余地が多分にあるということである。それぞれの解説者が不十分な点が多くあることを認めながらも、その可能性に期待している。深層学習は、今後、機械学習あるいは人工知能の分野において、注目すべき重要な技術であろう。そして、その技術的な可能性を無条件に期待することなく、あるいは、過去の失敗を理由に無条件に卑下することなく、正しく見極めることが大切である。本書がその一助になれば幸いである。

最後に、本書の内容について貴重な意見を下さった以下の方々に感謝したい：海野裕也、大塚誠、大野健太、小林颯介、坪井祐太、中山英樹、福島邦彦、渡辺太郎、渡部晋治。

2015年9月
松尾 豊
松原 仁

執筆者一覧

第 1 章 麻生 英樹

第 2 章 安田 宗樹

第 3 章 前田 新一

第 4 章 岡野原 大輔

第 5 章 岡谷 貴之

第 6 章 久保 陽太郎

第 7 章 ボレガラ ダヌシカ

編 集 神鳶 敏弘

目 次

序文	i
数式の表記	xi
深層学習手法の全体像	xiii
第Ⅰ部 基礎編 1	
第1章 階層型ニューラルネットワークによる深層学習 3	
1.1 はじめに	3
1.2 内部表現のデータからの学習	5
1.2.1 内部表現の重要性とその学習法	5
1.2.2 特徴工学と表現学習	7
1.3 階層型ニューラルネットワーク	8
1.3.1 ニューラルネットワーク研究の系譜	9
1.3.2 階層型ニューラルネットワークの数理モデル	9
1.4 階層型ニューラルネットワークの学習	12
1.4.1 誤り訂正学習	12
1.4.2 誤差逆伝播学習	14
1.4.3 競合学習	15
1.5 深層ニューラルネットワークによる深層表現の学習	16
1.5.1 誤差逆伝播学習による内部表現の学習	16
1.5.2 深層ニューラルネットワークの学習	17
1.6 置み込みニューラルネットワーク	19
1.7 自己符号化器	22
1.7.1 自己符号化器とその学習	22
1.7.2 積層自己符号化器	23

1.8	1.7.3 スパース自己符号化器	24
	1.7.4 雑音除去自己符号化器	25
1.8	おわりに	25
	参考文献	27
第2章 深層ボルツマンマシン		31
2.1	はじめに	31
2.2	統計的機械学習の考え方——データ生成モデルの再現	32
2.3	マルコフ確率場とボルツマンマシン	35
	2.3.1 マルコフ確率場	35
	2.3.2 ボルツマンマシン	37
	2.3.3 ボルツマンマシンとホップフィールド・ネットワーク の関係	39
	2.3.4 ボルツマンマシンの学習へ	40
2.4	可視変数のみのボルツマンマシン学習	40
	2.4.1 カルバック-ライブラー・ダイバージェンスからの学 習方程式の導出	43
	2.4.2 ボルツマンマシン学習の実装と組み合わせ爆発の問題	44
2.5	隠れ変数ありのボルツマンマシン学習	46
	2.5.1 隠れ変数ありの場合の学習について	49
	2.5.2 隠れ変数を導入する意味	50
2.6	ボルツマンマシン上で近似手法	52
	2.6.1 ギブスサンプリング	52
	2.6.2 平均場近似	55
2.7	制限ボルツマンマシン	57
	2.7.1 条件付き独立性に基づく制限ボルツマンマシンの性質	59
	2.7.2 制限ボルツマンマシンの学習	60
2.8	深層ボルツマンマシン	62
	2.8.1 深層ボルツマンマシンの事前学習	64
	2.8.2 事前学習後の最尤推定法に基づく学習	67
	2.8.3 自己符号化器としての制限ボルツマンマシン	70
	2.8.4 深層ボルツマンマシンの利用法	71
2.9	深層信念ネットワーク	72
	2.9.1 深層信念ネットワークに対する事前学習と推論	74

2.9.2	深層信念ネットワークに対する事前学習の正当性	75
2.10	おわりに	80
	参考文献	80
第3章	事前学習とその周辺	83
3.1	はじめに	83
3.2	自由度の高い統計モデルの学習における困難とその解決法	84
3.2.1	学習を難しくする要因	84
3.2.2	既存の解決法	85
3.2.3	新たな解決法	88
3.3	自己符号化器による内部表現の学習	91
3.3.1	自己符号化器とその損失関数の定義	91
3.3.2	層ごとの貪欲学習を用いた自己符号化器の事前学習	93
3.4	確率的なモデルを用いた事前学習	93
3.4.1	制限ボルツマンマシン	94
3.4.2	指数型ハーモニウム族	96
3.4.3	指数型ハーモニウム族のコントラスティブ・ダイバージェンス法による学習	100
3.4.4	コントラスティブ・ダイバージェンス法が最適化している損失関数	100
3.4.5	コントラスティブ・ダイバージェンス法と類似した学習則を与えるアルゴリズム	107
3.4.6	コントラスティブ・ダイバージェンス法から派生した学習則	109
3.4.7	確率的なモデルの事前学習と自己符号化器の学習の関係	111
3.5	確定的なモデルを用いた事前学習	112
3.5.1	教師なし学習による確定的なモデルの学習	113
3.5.2	教師あり学習による確定的なモデルの学習	117
3.6	Product of Experts の学習法としてのコントラスティブ・ダイバージェンス法	118
3.7	おわりに	119
	参考文献	120

第4章 大規模深層学習の実現技術	125
4.1 はじめに	125
4.2 深層学習の最適化	126
4.2.1 深層学習の基本計算	127
4.2.2 確率的勾配降下法	128
4.3 高速化手法	129
4.3.1 分散並列処理: DistBelief	130
4.3.2 GPU を用いた大規模ニューラルネットの実現	133
4.3.3 InfiniBand の利用	135
4.3.4 学習収束の高速化	136
4.4 過学習制御: DropOut	139
4.5 活性化関数	142
4.5.1 ReLU	143
4.5.2 MaxOut	143
4.6 学習率の調整	144
4.6.1 AdaGrad	144
4.6.2 Adam	145
4.6.3 超パラメータの最適化	146
4.7 実装技術	147
4.7.1 実装の正しさのチェック	147
4.8 おわりに	148
参考文献	148
第II部 応用編	151
第5章 画像認識のための深層学習	153
5.1 はじめに	153
5.1.1 畳み込みニューラルネットワークの再発見	154
5.1.2 その後の研究	154
5.2 畳み込みニューラルネットワーク	156
5.2.1 基本構造	156
5.2.2 畳み込み層	157
5.2.3 プーリング層	160

5.2.4	例：手書き数字認識のための畳み込みニューラルネット	161
5.2.5	学習	163
5.2.6	コントラスト調整とデータの正規化	164
5.3	畳み込みニューラルネットワークのはたらき	166
5.3.1	一般物体認識の難しさ	166
5.3.2	一般物体認識の従来法	167
5.3.3	従来法と畳み込みニューラルネットワークの比較	170
5.3.4	ネットワークの構造と認識性能	173
5.3.5	畳み込みニューラルネットワークの拡張の試み	174
5.4	畳み込みニューラルネットワークの内部表現	175
5.4.1	可視化	175
5.4.2	脳神経系との関係	176
5.4.3	転移学習	176
5.5	画像特徴の教師なし学習	178
5.5.1	単層自己符号化器による局所特徴の学習	178
5.5.2	多層ネットワークによる特徴学習	181
5.6	おわりに	184
	参考文献	185
第6章	音声認識のための深層学習	189
6.1	はじめに	189
6.2	音声認識	190
6.2.1	音声認識に用いられるモデル	191
6.2.2	大語彙連続音声認識システムの構成	194
6.3	音声認識におけるニューラルネットワーク	195
6.3.1	時間遅れニューラルネットワーク	197
6.3.2	隠れマルコフモデルと組み合わせたニューラルネット：ハイブリッド方式	198
6.3.3	隠れマルコフモデルと組み合わせたニューラルネット：タンデム方式	200
6.4	音響モデルにおける深層学習：事前学習	201
6.4.1	制限ボルツマンマシンによる事前学習を用いた深層ニューラルネットワーク-隠れマルコフモデル	202
6.4.2	雑音除去自己符号化器による事前学習	204

6.4.3 識別的事前学習	205
6.5 音響モデルにおける深層学習：学習とモデルの進展	206
6.5.1 系列識別学習	207
6.5.2 回帰結合ニューラルネットワークによる音響モデル . .	211
6.5.3 Long Short-Term Memory 法	214
6.5.4 マルチストリーム・マルチタスク学習	217
6.6 言語モデルにおける深層学習	218
6.6.1 回帰結合ニューラルネットワークによる言語モデル . .	218
6.7 おわりに	220
参考文献	222
第7章 自然言語処理のための深層学習	225
7.1 はじめに	225
7.2 深層学習と言語モデル	228
7.2.1 ニューラルネットワーク言語モデル	229
7.2.2 その他の言語モデル	231
7.3 単語の意味表現学習	233
7.3.1 ボトムアップ的な意味表現構築手法	233
7.3.2 トップダウン的な意味表現予測手法	234
7.3.3 階層型ソフトマックスによる計算	239
7.3.4 意味表現学習手法のその他の話題	240
7.4 深層学習と意味構築	243
7.4.1 言い換え表現認識への応用	244
7.5 おわりに	248
参考文献	249
和文索引	253
英文索引	261

数式の表記

本書で用いる表記をここにまとめておく。

変数や行列

大小文字・書体により、ベクトル・スカラーや、通常変数・実現値・確率変数・集合を次のように区別する。

大小文字・書体による変数の種類の区分

表記	内容
x (小文字イタリック体)	スカラー変数、スカラー実現値
X (大文字イタリック体)	スカラー確率変数
\mathbf{x} (小文字ボールドローマン体)	ベクトル変数、確率変数のベクトル実現値
\mathbf{X} (大文字ボールドローマン体)	行列変数、ベクトル確率変数
X (大文字カリグラフィック体)	集合変数

なお、確率変数が特定の実現値をとる場合の確率 $p(X = x)$ や $p(\mathbf{X} = \mathbf{x})$ は、特に混乱が生じない場合は $p(x)$ や $p(\mathbf{x})$ のように略記する。

下記の変数については、本書を通じて同じ内容を表すものとする。

変数の表す内容

変数	内容
$x, \mathbf{x}, \mathbf{X}$	入力信号・情報
$y, \mathbf{y}, \mathbf{Y}$	出力信号・情報
$w, \mathbf{w}, \mathbf{W}$	結合重み
$v, \mathbf{v}, \mathbf{V}$	可視変数（観測変数）
$h, \mathbf{h}, \mathbf{H}$	隠れ変数（潜在変数）
$\theta, \boldsymbol{\theta}, \Theta$	パラメータ
\mathbf{I}	単位行列
$\mathbb{R}, \mathbb{N}, \mathbb{Z}$	それぞれ実数、自然数、整数の集合
$\mathbf{0}$	0ベクトル（要素がすべて0であるベクトル）

演算と関数

本書を通じて用いる演算と関数の表記を以下にまとめておく。

演算と関数の共通表記

変数	内容
\mathbf{X}^T	行列 \mathbf{X} の転置
\mathbf{X}^{-1}	行列 \mathbf{X} の逆行列
$\mathbf{X} \circ \mathbf{Y}$	行列の要素ごとの積
$f * g$	畳み込み演算
$\text{Dom}(x)$	変数 x の定義域
$\text{diag}(\mathbf{d})$	ベクトル \mathbf{d} を対角要素とする対角行列
$H(p)$	確率分布 p のエントロピー関数: $-\sum p \log p, -\int p \log p$
$a(x)$	活性化関数: ニューラルネットワークのノード内で、重みと入力の内積を出力に変換する関数
$\text{sig}(x)$	シグモイド関数: $1/(1 + e^{-x})$

ベクトル引数の関数 スカラー関数 $f(x)$ に対して、その引数がベクトルである表記 $f(\mathbf{x})$ は、ベクトル \mathbf{x} の各要素を関数 f に適用して得られるベクトルを表す。

確率質量関数と確率密度関数 確率変数 X が離散の場合の確率質量関数も、連続値の場合の確率密度関数も特に区別することなく $p(X)$ と表記する。

期待値 $E_{p(X)}[f(X)]$ は、分布 $p(X)$ についての次の期待値を表す：

$$\sum_{x \in \text{Dom}(X)} f(x)p(x) \quad — X \text{ が離散の場合}$$

$$\int_{x \in \text{Dom}(X)} f(x)p(x)dx \quad — X \text{ が連続の場合}$$

なお、 $p(X)$ を省略した場合は、関数 f のすべての確率変数の同時分布に関する期待値を表す。例えば、 $E[f(X, Y)]$ は、 $E_{p(X, Y)}[f(X, Y)]$ の意味である。

カルバック-ライブラーダイバージェンス 確率分布 $p(X)$ と $q(X)$ の間のカルバック-ライブラーダイバージェンスを次のように表記する：

$$D_{\text{KL}}(p(X) \parallel q(X)) = \int_{x \in \text{Dom}(X)} p(x) \log \frac{p(x)}{q(x)} dx$$

深層学習手法の全体像

本書では様々なニューラルネットワークのモデルを紹介するが、深層学習とそれ以外の手法という点と、確定的・確率的モデルかという二つの観点から、これらのモデルを関連づけてその全体像を示しておく。その後、確定的と確率的モデルに分けてそれぞれの手法について簡単に紹介する。

まず、カラーページ最初の『深層学習手法の全体像』の左側の確定的モデルでは、入力に対して確定的にその出力が決まる。これらのモデルは、結合重みとバイアスをパラメータとする線形変換と、この変換後の値を非線形変換する活性化関数で決まるノードで構成されていることが多い。図の右側にまとめたもう一方の確率的モデルは入出力変数の同時分布を表す。データ集合に含まれる入出力値からその値がわかる観測変数と、値がわからない隠れ変数があり、これらの変数間の依存関係をグラフィカルモデルという表現方法で記述している。確定的と確率的のどちらにも、2~3層の浅いのニューラルネットワークと、それ以上の層数で構成される深層ニューラルネットワークとがある。

なお、読者が深層学習の全体を概観するのに参考となるように各種のモデルを分類した図を示したが、この分類はあくまで、非深層・深層と、確定的・確率的モデルという特定の観点からの分類にすぎないことに留意されたい。例えば、確率的モデルの制限ボルツマンマシンは、確定的モデルの深層自己符号化器の事前学習に利用されている。このことから、ニューラルネットワークの各手法の関係は複雑であり、いろいろな分類が可能なことが想像できるだろう。

確定的モデル

確定的ニューラルネットワークの多くは階層型ニューラルネットワークと自己符号化器である。

階層型ニューラルネットワーク型

階層型ニューラルネットワークは、入力から出力へ結合を通じて信号が順伝播するフィードフォワード型の構造をしており、主に教師あり学習に用いられる。この種のニューラルネットワークは図の最も左にまとめてあり、これらを順に列挙する。

パーセptron 【1.4.1 項】

2層でのみ構成される最初に提案されたニューラルネットであり、図では階層型ニューラルネットの一種である。線形分離可能と呼ばれる条件を満たす問題しか解けない制限があるが、解ける問題については誤り訂正学習則により有限回の更新で学習が収束することが証明されている。

階層型ニューラルネット 【1.3 節】

深層学習の登場前に最も使われていたのは、3層前後のフィードフォワード構造を備えた階層型ニューラルネットである。このモデルは、多層パーセptronとも呼ばれ、誤差逆伝播学習という効率的な学習法が開発されたことにより1980年代に普及した。教師あり学習に用いる場合がほとんどだが、競合学習則を適用して教師なし学習に用いる場合もある。

深層（階層型）ニューラルネット 【1.5 節】

深層ニューラルネットは、広義には多層のニューラルネットワーク全般のことを目指し、狭義にはフィードフォワード構造の階層型ニューラルネットを4層以上に拡張したものを目指す。この狭義のニューラルネットは、3層でも中間層のノード数が十分であれば任意の関数を近似できるという理論的利点を備えていたが、局所最適解や勾配消失問題などの技術的問題のため、広く使われることはなかった。しかし、中間ノード数を増加させるより、深層化のほうが効果的に予測性能を向上できることが発見的に示されたことに加えて、事前学習やDropOutなどの新技術の登場や、活性化関数やネットワーク構造などの工夫により技術的問題に対処できるようになったことから2010年代に普及した。

再帰ニューラルネット 【7.4.1 項】

再帰的部分構造を組み込んだニューラルネットであり、再帰的な構造を備えた内部表現を獲得する目的で用いる。下位の部分木構造から、上位の部分木構造を再帰的に構成する木構造になっているため、再帰ニューラルネットワークと呼ばれている。図では階層型ニューラルネットに含めているが、本書では教師なし学習を行う自己符号化器に再帰的構造を組み込んだ再帰自己

符号化器を取り上げる。

畳み込みニューラルネット [5.2 節]

畳み込み構造を組み込んだ階層型ニューラルネットワークである。深層学習の登場以前からネオコグニトロンや LeNet などとして提案されていた。2010 年代の分散並列計算技術の進展と、学習用データの大規模化により、画像認識の分野で特に普及している。

回帰結合ニューラルネット [6.5.2 項]

回帰結合ニューラルネットは、系列データを処理する目的で考案された。そのため、前回の時刻の入力の情報を、現在の入力の処理に伝えるための回帰結合入力を備えている。畳み込みニューラルネットと同様に、深層学習の登場以前から提案されていたが、2010 年代に学習の大規模化に伴って、音声認識や自然言語処理分野で普及している。勾配消失問題に対処した long short-term memory 法などの改良もなされている。

自己符号化器型

もう一方の確定的モデルである自己符号化器は、砂時計型のニューラルネットワークであり、教師なし学習を行う。自己符号化器は図の中央にまとめてあり、これらを順に列挙する。

自己符号化器 [3.3.1 項, 1.7 節]

基本型の自己符号化器は、教師なし学習により入力の低次元表現を獲得する目的で考案された。入力を中間層で低次元表現に変換する符号化と、この低次元表現を元の次元の表現に戻す復号化とを組み合わせた、3 層の砂時計型の構造をしている。入力信号と出力信号の間の再構成誤差を小さくするように学習を行う。

深層自己符号化器 [3.3.1 項]

深層自己符号化器は、入力から中間層までの符号化部分と、中間層から出力層までの復号化部分の層数を増やして深層化した自己符号化器である。

積層自己符号化器 [1.7.2 項]

積層自己符号化器は、層ごとの貪欲学習により深層自己符号化器の勾配消失問題を回避したものである。

雑音除去自己符号化器 [3.5.1 項, 6.4.2 項]

雑音除去自己符号化器は、入力信号に雑音を加えることで、未知の信号に対するロバスト性を向上させる手法である。

確率的モデル

ニューラルネットワークの分野で現在広く使われている確率モデルはボルツマンマシンに由来するものである。図の右端にまとめたこれらのモデルを順に紹介する。

ボルツマンマシン【2.3 節】

ボルツマンマシンは、マルコフ確率場という確率モデルの一種である。観測・隠れ変数を各ノードとし、これらのノード間の依存関係を無向の結合で示したグラフィカルモデルで記述する。組み合わせ爆発問題により学習が困難なため、あまり利用されていない。

制限ボルツマンマシン【2.7 節, 3.4.1 項】

制限ボルツマンマシンは、ボルツマンマシンに、観測変数と隠れ変数の間にしか依存関係がないように制限を加えた確率モデルである。この制限で格段に学習が効率化され、実用的な問題に適用されるようになった。ハーモニウムという名称で提案されたが、現在では制限ボルツマンマシンの呼称が定着している。観測変数と隠れ変数をそれぞれ1層と見なすと、全体で2層で構成されている。このモデルを指数族に一般化したものを指数型ハーモニウム族と呼ぶ。

深層ボルツマンマシン【2.8 節】

深層ボルツマンマシンは、制限ボルツマンマシンの隠れ変数の層をそのまま多段にすることで深層化したモデルである。事前学習やコントラスティブ・ダイバージェンス法などの手法の開発が進み、2010年代に広く利用されるようになった。

深層信念ネットワーク【2.9 節】

深層信念ネットワークは、隠れ変数を多層にして深層化する点では深層ボルツマンマシンと同じだが、その依存関係を無向の結合ではなく、有向の結合によって表現したものが深層信念ネットワークである。事前学習などの技術が活用でき、2010年代に普及が進んでいる。

和文索引

【記号・英数字】

1-of- n 表現 (1-of- n representation)	133, 163, 229	EFH	指数型ハーモニウム族
Adagrad	144	ELM	極端学習機械
Adam	145	FA	因子分析
ASO	交互構造最適化	GMM	混合正規分布
backpropagation through time 法	128, 212	GPU	133
bag-of-features モデル (bag-of-features model)		HMM	隠れマルコフモデル
169		HNN	階層型ニューラルネットワーク
bag-of-words モデル (bag-of-words model)	226, 236	ICA	独立成分分析
bag-of-visual-words モデル	bag-of-features モデル	ILSVRC	153
BM	ボルツマンマシン	InfiniBand	135
BoF モデル	bag-of-features モデル	KL 距離 (KL distance)	カルバック-ライブ ラー・ダイバージェンス
BoW モデル	bag-of-words モデル	KL 情報量 (KL information)	カルバック-ラ イブラー・ダイバージェンス
BPTT 法	backpropagation through time 法	KL ダイバージェンス (KL divergence)	カル バック-ライブルー・ダイバージェンス
CDBN	畳み込み深層信念ネットワーク	L-BFGS 法 (L-BFGS method)	131, 246
CD 法	コントラスティブ・ダイバージェン ス法	L_1 正則化 (L_1 regularization)	25
CNN	畳み込みニューラルネットワーク	lasso	L_1 正則化
ConvNet	畳み込みニューラルネットワーク	LCA	局所コントラスト正規化
DAE	深層自己符号化器	LeNet	畳み込みニューラルネットワーク
DBL 法	detailed balance learning 法	limited-memory Broyden-Fletcher-Goldfarb- Shanno 法	L-BFGS 法
DBM	深層ボルツマンマシン	long short-term memory 法	214
DBN	深層信念ネットワーク	low effective dimension	146
detailed balance learning 法 (detailed balance learning method)	104	L_p プーリング (L_p pooling)	160
DistBelief	130	LSTM 法	long short-term memory 法
DNN	深層ニューラルネットワーク	MAP 推定 (MAP estimation)	最大事後確率 推定
DNN-HMM	深層ニューラルネットワーク－ 隠れマルコフモデル	MaxOut	143
Downpour SGD	131	MFCC	メル周波数ケプストラム係数
DropOut	139	MLE	最尤推定
		MMI 標準 (MMI criterion)	最大相互情報量 規準

MPF 法	☞ 最小確率流法	192
MRF	☞ マルコフ確率場	110
MV-RNN	☞ matrix vector recursive neural network	13
NLMM	☞ ニューラルネットワーク言語モデル	
NN	☞ ニューラルネットワーク	
NN-HMM	☞ ニューラルネットワーク-隠れマルコフモデル	
<i>n</i> グラム (<i>n</i> -gram)	194, 226, 228	
PCA	☞ 主成分分析	
PoE	☞ product of experts	
product of experts	118	
pyramid match kernel	170	
RBM	☞ 制限ボルツマンマシン	
rectified linear unit	ReLU	
ReLU	143, 156	
RNN	☞ 回帰結合ニューラルネットワーク	
Sandblaster L-BFGS	131	
scale invariant feature transform	☞ SIFT	
SGD	☞ 確率的勾配降下法	
SIFT	168	
TDNN	☞ 時間遅れニューラルネットワーク	
TICA	☞ トポグラフィック独立成分分析	
word2vec	235	
【あ】		
誤り訂正学習 (error correction learning)	12	
暗黒知識 (dark knowledge)	138	
鞍点 (saddle point)	85	
言い換え表現認識 (paraphrase detection)	244	
一対他符号化 (one-versus-rest encoding)	☞ 1-of- <i>n</i> 表現	
一般物体認識 (general object recognition)	153, 166	
～の従来法	167	
意味表現 (semantic representation)	232, 233	
因子分析 (factor analysis)	8	
エネルギー関数 (energy function)	36	
エルマン・ネットワーク (Elman network)	211	
おばあさん細胞 (grandmother cell)	153	
重み (weight)	10	
重み共有 (weight sharing)	159	
重み減衰 (weight decay)	164	
音響モデル (acoustic model)	191	
音声認識 (speech recognition)	190	
音素 (phoneme)	189	
音素文脈 (phonemic context)	192	
温度パラメータ (temperature parameter)	110	
オンライン学習 (online learning)	13	
【か】		
回帰結合ニューラルネットワーク (recurrent neural network)	211	
long short-term memory 法	214	
エルマン・ネットワーク	211	
音響モデル	211	
言語モデル	218	
双方向型～	211	
階層型ソフトマックス (hierarchical softmax)	239	
階層型ニューラルネットワーク (hierarchical neural network)	8, 127, 196	
～の学習	12	
誤差逆伝播法	14	
深層～	☞ 深層ニューラルネットワーク	
～モデル	9	
ガウス-ベルヌーイ型制限ボルツマンマシン (Gaussian-Bernoulli restricted Boltzmann machine)	202	
ガウス雑音 (Gaussian noise)	25	
過学習 (over-fitting)	84, 139	
係り受け解析 (dependency parsing)	225	
学習率 (learning rate)	12, 164	
～の調整	144	
確率的グラフィカルモデル (probabilistic graphical model)	☞ グラフィカルモデル	
確率的勾配降下法 (stochastic gradient descent method)	14, 128	
確率的最大プーリング (probabilistic max pooling)	181	
隠れ層 (hidden layer)	11, 58	
隠れ変数 (hidden variable)	40, 50, 94	
隠れマルコフモデル (hidden Markov model)	192	
深層ニューラルネットワーク～	201	
ニューラルネットワーク～	195	
可視層 (visible layer)	58	
可視変数 (visible variable)	40, 94	
画像認識 (image recognition)	153	
活性化関数 (activation function)	10, 142, 159	
ReLU～	143	
恒等～	22	

しきい～	10	交差エントロピー (cross entropy)	163, 199
シグモイド～	14	恒等関数 (identity function)	22
ソフトマックス～	157	勾配降下法 (gradient descent method)	45, 127
ガボールフィルタ (Gabor filter)	171	勾配消失問題 (vanishing gradient problem)	18, 214
カルバック-ライブラー・ダイバージェンス (Kullback-Leibler divergence)	43, 97	勾配上昇法 (gradient ascent method) \Rightarrow 勾配降下法	
頑健性 (robustness)	\Rightarrow ロバスト性	コーパス (corpus)	228
観測変数 (observable variable)	\Rightarrow 可視変数	誤差逆伝播法 (backpropagation method)	14, 127
機械学習 (machine learning)	4	内部表現学習	16
規格化定数 (normalization constant)	37, 94	コスト関数 (cost function) \Rightarrow 損失関数	
擬似負例 (pseudo negative instance)	242	ごま塩雜音 (salt-and-pepper noise)	25
機能語 (function word)	242	混合正規分布 (Gaussian mixture model)	192
ギブスサンプラー (Gibbs sampler) \Rightarrow ギブスサンプリング		コントラスティブ・ダイバージェンス (contrastive divergence)	103
ギブスサンプリング (Gibbs sampling)	52	コントラスティブ・ダイバージェンス法 (contrastive divergence method)	99
ギブス分布 (Gibbs distribution) \Rightarrow ボルツマン分布		継続的～	109
逆畳み込みネットワーク (deconvolutional network)	175	損失関数	100
競合学習 (competitive learning)	15		
教師あり学習 (supervised learning)	12		
教師なし学習 (unsupervised learning)	12		
共有重み (shared weight)	159		
局所コントラスト正規化 (local contrast normalization)	164		
局所最適解 (local optimum)	15, 85		
局所特徴 (local feature)	167		
極端学習機械 (extreme learning machine)	116		
組み合わせ爆発 (combinatorial explosion)	45		
グラフィカルモデル (graphical model)	37		
グリッド探索 (grid search)	146		
経験分布 (empirical distribution)	43		
継続的コントラスティブ・ダイバージェンス法 (persistent contrastive divergence method)	109		
系列識別学習 (sequence discriminative training)	207		
欠落雜音 (masking noise)	25		
言語モデル (language model)	192, 228		
ニューラルネットワーク～	229		
減算正規化 (subtractive normalization)	164		
交換モンテカルロ法 (exchange Monte Carlo method)	110		
交互構造最適化 (alternating structural optimization)	227		
交互最適化 (alternating optimization)	238		
【さ】			
再帰自己符号化器 (recursive autoencoder)	244		
再帰ニューラルネットワーク (recursive neural network)			
再帰自己符号化器	244		
展開再帰自己符号化器	246		
最急勾配降下法 (steepest gradient descent method)		\Rightarrow 勾配降下法	
再構成型 TICA (reconstruction TICA) \Rightarrow 再構成型トポグラフィック独立成分分析			
再構成型トポグラフィック独立成分分析 (reconstruction topographic independent component analysis)			182
再構成誤差 (reconstruction error)	92		
最小確率流法 (minimum probability flow method)			108
最大事後確率推定 (maximum a posteriori estimation)			34
最大相互情報量規準 (maximum mutual information criterion)			208
最大プーリング (max pooling)			160
最尤推定 (maximum likelihood estimation)	41		
最尤推定量 (maximum likelihood estimator)	41		
雑音除去自己符号化器 (denoising autoencoder)			
25, 115, 204			

- 時間遅れニューラルネットワーク (time-delay neural network) 197
 しきい関数 (threshold function) 10
 識別的事前学習 (discriminative pre-training)
 205
 シグモイド関数 (sigmoid function) 14
 シグモイド信念 (sigmoid belief) 39
 次元削減 (dimension reduction) 5, 8
 事後学習 (post-training) 227
 自己符号化器 (autoencoder)
 22, 91, 178
 ～の学習 22
 再帰～ 244
 雑音除去～ 204
 ～の事前学習 93
 周辺化雑音除去～ 115
 縮小～ 25
 深層～ 92
 スペース～ 24
 ～としての制限ボルツマンマシン 70
 積層～ 70
 展開再帰～ 246
 変分～ 112
 事後分布 (posterior distribution) 34
 指数型ハーモニウム族 (exponential family har-
 monium) 96
 ギブスサンプリング 98
 事前学習 (pre-training) 23, 88, 112, 226
 確率的モデルによる～ 93
 雑音除去自己符号化器の～ 204
 識別的～ 205
 自己符号化器の～ 112
 深層信念ネットワークの～ 74
 深層ボルツマンマシンの～ 64
 積層自己符号化器 113
 自然言語処理 (natural language processing) 225
 自然勾配法 (natural gradient method) 88
 事前分布 (prior distribution) 34
 シナプス (synapse) 9
 周辺化雑音除去自己符号化器 (marginalized denoising autoencoder) 115
 縮小自己符号化器 (contractive autoencoder) 25
 主成分分析 (principal component analysis) 8
 出力関数 (output function) 8 活性化関数
 出力層 (output layer) 11
 受容野 (receptive field) 11, 159
 条件付き独立性 (conditional independence) 59
 詳細釣り合い条件 (detailed balance condition) 105
 蒸留 (distillation) 137
 除算正規化 (divisive normalization) 165
 人工知能 (artificial intelligence) 3
 深層階層型ニューラルネットワーク (deep hierarchical neural network) 8 深層ニューラルネットワーク
 深層学習 (deep learning) 3
 深層自己符号化器 (deep autoencoder) 92
 深層信念ネットワーク (deep belief network) 72
 ～の事前学習 74
 ～の推論 74
 畳み込み～ 181
 深層ニューラルネットワーク (deep neural net-
 work) 16
 音響モデル 206
 ～の学習 17
 ～-隠れマルコフモデル 201
 言語モデル 218
 ～の事前学習 202
 深層ニューラルネットワーク-隠れマルコフ
 モデル (deep neural network-hidden Markov
 model) 201
 深層ボルツマンマシン (deep Boltzmann ma-
 chine) 62
 ～の学習 67
 ～の事前学習 64
 ～の利用法 71
 推定 (estimation) 35
 推論 (inference) 35
 砂時計型ニューラルネットワーク (hourglass-type neural network) 22, 91
 スペース自己符号化器 (sparse autoencoder) 24, 179
 スパース性 (sparseness) 7, 8
 正規化定数 (normalization constant) 8 規格化
 定数
 制限ボルツマンマシン (restricted Boltzmann
 machine) 57, 94, 202
 ガウス-ベルヌーイ型～ 202
 ～の学習 60
 自己符号化器としての～ 70
 ～の条件付き独立性 59
 ベルヌーイ-ベルヌーイ型～ 57

生成モデル (generative model)	33	～の学習	163
正則化 (regularization)	24	～の拡張	174
DropOut	142	従来法との比較	170
L_1 ～	25	全結合層	157
蒸留	138	畳み込み層	157
スペース～	179	内部表現	175
積層自己符号化器 (stacked autoencoder)	23, 70,	認識性能	173
113		プリーリング層	160
ゼロ頻度問題 (zero-frequency problem)	194,	マルチタスク学習	176
228		多様体学習 (manifold learning)	8
線形分離 (linear separation)	12	単語エラー率 (word error rate)	207
線形分離可能 (linearly separable)	13	単語集合モデル (bag-of-words model) \bowtie bag-of-words モデル	
全結合層 (fully connected layer)	157	単語袋詰めモデル (bag-of-words model) \bowtie bag-of-words モデル	
潜在表現 (latent representation)	\bowtie 内部表現	単純パーセプトロン (simple perceptron)	12
潜在変数 (latent variable)	\bowtie 隠れ変数	中間層 (internal layer) \bowtie 隠れ層	
相互結合ニューラルネットワーク (mutually connected neural network)	11	中心差分 (central difference)	147
層ごとの貪欲学習 (greedy layer-wise training)	93, 112	超パラメータ (hyper parameter)	146
双方向型回帰結合ニューラルネットワーク (bi-directional recurrent neural network)	211	通時的誤差逆伝播法 \bowtie backpropagation through time 法	
疎性 (sparseness)	\bowtie スパース性	データ並列化 (data parallelism)	130
ソフトマックス関数 (softmax function)	157	転移学習 (transfer learning) \bowtie マルチタスク学習	
～出力層	198	展開再帰自己符号化器 (unfolding recursive autoencoder)	246
損失関数 (loss function)	84, 127	伝承サンプリング (ancestral sampling)	75
【た】			
大域特徴 (global feature)	167	統計的機械学習 (statistical machine learning)	33
大語彙連続音声認識 (large vocabulary continuous speech recognition)	194	動的プリーリング (dynamic pooling)	247
対数双線形 (log-bilinear form)	238	特徴 (feature)	5
対数メルフィルタバンク特徴ベクトル (log Mel-filterbank feature vector)	203	特徴学習 (feature learning) \bowtie 表現学習	
対数尤度関数 (log-likelihood function)	41	特徴工学 (feature engineering)	7
多層ニューラルネットワーク (multi-layer neural network) \bowtie 階層型ニューラルネットワーク	10	特徴点 (feature point)	65
多層パーセプトロン (multi-layer perceptron) \bowtie 階層型ニューラルネットワーク	10	独立成分分析 (independent component analysis)	
畳み込み (convolution)	157	8	
畳み込み深層信念ネットワーク (convolutional deep belief network)	181	トポグラフィック・マッピング (topographic mapping)	183
畳み込み層 (convolution layer)	157	トポグラフィック独立成分分析 (topographic independent component analysis)	182
畳み込みニューラルネットワーク (convolutional neural network)	19,	再構成型～	182
156		トライグラム (trigram) \bowtie n グラム	
【な】			
内部共変量シフト (internal covariate shift)	136		
内部表現 (internal representation)	5, 16		
二分構文木 (binary parse tree)	244		

- ニューラルネットワーク (neural network) 3
 回帰結合～ 211
 階層型～ 8
 ～-隠れマルコフモデル 195
 時間遅れ～ 197
 深層～ 16
 砂時計型～ 22
 相互結合～ 11
 多層～ \Rightarrow 階層型ニューラルネットワーク
 置込み～ 19, 156
 フィードフォワード型～ \Rightarrow 階層型ニューラルネットワーク
 ランダムネットワーク
 ～の歴史 9
 ニューラルネットワーク-隠れマルコフモデル
 (neural network-hidden Markov model) 195
 タンデム方式 200
 ハイブリッド方式 198
 ニューラルネットワーク言語モデル (neural network language model) 229
 入力層 (input layer) 11
 ニューロン (neuron) 9
 ネオコグニトロン (neocognitron) 19
 ノイズ (noise) \Rightarrow 雑音
 脳神経系 (cerebral nerve system) 9, 176
 ノーフリーランチ定理 (no free lunch theorem) 6
- 【は】**
- パーセプトロン (perceptron) 9
 ～学習 \Rightarrow 誤り訂正学習
 ～学習則の収束定理 13
 ～の限界 13
 多層～ \Rightarrow 階層型ニューラルネットワーク
 パーセプトロン学習 (perceptron learning) \Rightarrow 誤り訂正学習
 ハーモニウム (harmonium) \Rightarrow 制限ボルツマンマシン
 バイグラム (bigram) $\Rightarrow n$ グラム
 白色化 (whitening) 165
 発音辞書モデル (pronunciation model) 192
 バックオフ平滑化 (back-off smoothing) 229
 バッチ学習 (batch learning) 13
 バッチ正規化 (batch normalization) 136
 ハフマン木 (Huffman tree) 239
 パラレル・テンパリング法 (parallel tempering method) \Rightarrow 交換モンテカルロ法
- 反復的パラメータ混合法 (iterative parameter mixing method) 242
 ビジュアルワード (visual word) 169
 表現学習 (representation learning) 7, 178, 226
 ヒンジ損失 (hinge loss) 232
 品詞タグ付け (part-of-speech tagging) 225
 フィードフォワード型ニューラルネットワーク (feed-forward neural network) \Rightarrow 階層型ニューラルネットワーク
 フィッシャーベクトル (Fisher vector) 170
 プーリング (pooling) 21, 160, 170
 L_p ～ 160
 確率的最大～ 181
 最大～ 160
 動的～ 247
 平均～ 160
 プーリング層 (pooling layer) 160
 復号化器 (decoder) 92
 符号化器 (encoder) 92
 不変性 (invariance) 7, 89, 160, 174, 183, 238
 不要語 (stop word) 242
 プラトー (plateau) 85
 負例サンプリング (negative sampling) 241
 ブロック化ギブスサンプリング (blocked Gibbs sampling) 55
 分散的意味表現 (distributed semantic representation) 234
 分散並列計算 (distributed parallel computation) 130
 分子場近似 (molecular field approximation)
 \Rightarrow 平均場近似
 分配関数 (partition function) \Rightarrow 規格化定数
 分布意味論 (distributional semantics) 229
 分布仮説 (distributional hypothesis) 233
 分布的意味構築 (distributional semantic composition) 243
 分布的意味表現 (distributional semantic representation) 234
 分布メモリ (distributional memory) 244
 文脈 (context) 229
 平滑化 (smoothing) 229
 平均場近似 (mean-field approximation) 55
 平均場方程式 (mean-field equation) 57
 平均プーリング (average pooling) 160
 ヘビサイド関数 (Heaviside function) \Rightarrow しきい値関数

変分自己符号化器 (variational autoencoder)	112	モーメント・マッチング (moment matching)	42
変分ベイズ法 (variational Bayes method)	☞ 平均場近似	目的関数 (objective function)	☞ 損失関数
ホップフィールド・ネットワーク (Hopfield network)	39	モデル誤差 (model error)	51
ボルツマン分布 (Boltzmann distribution)	36	モデルバイアス (model bias)	☞ モデル誤差
ボルツマンマシン (Boltzmann machine)	37	モデル並列化 (model parallelism)	130
～の学習	40	モメンタム (momentum)	164
～の学習方程式	42		
～の近似手法	52		
深層～	62		
制限～	57		
<hr/>			
【ま】		【や】	
マルコフ確率場 (Markov random field)	35	尤度関数 (likelihood function)	41
マルコフ性 (Markov property)	53	ユニグラム (unigram)	☞ <i>n</i> グラム
マルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo method)	52		
マルチストリーム学習 (multi-stream learning)			
217			
マルチタスク学習 (multitask learning)	217, 233		
畳み込みニューラルネットワーク	176		
醜いアヒルの子の定理 (ugly duckling theorem)			
6			
ミニバッチ法 (mini batch method)	129		
無向グラフ (undirected graph)	36		
メル周波数ケプストラム係数 (mel-frequency cepstral coefficient)	193		
<hr/>			
【ら】		【ら】	
ランダム探索 (random search)	146		
リグレット (regret)	145		
類推問題 (analogical reasoning problem)	242		
連鎖律 (chain rule)	128		
連続 bag-of-words モデル (continuous bag-of-words model)	235		
連続スキップグラムモデル (continuous skip-gram model)	237		
ロス関数 (loss function)	☞ 損失関数		
ロバスト性 (robustness)	7		
<hr/>			
【わ】		【わ】	
ワイングラス型ニューラルネットワーク (wine-glass-type neural network)	☞ 砂時計型ニューラルネットワーク		
話者適応 (speaker adaptation)	200		

英文索引

【 Symbols / Numbers 】

1-of-*n* representation (1-of-*n* 表現) 133, 163,
229

【 A 】

acoustic model (音響モデル) 191

activation function (活性化関数) 10, 142, 159

identity –	22
ReLU –	143
sigmoid –	14
softmax –	157
threshold –	10

Adagrad 144

Adam 145

alternating optimization (交互最適化) 238

alternating structural optimization (交互構造最
適化) 227

analogical reasoning problem (類推問題) 242

ancestral sampling (伝承サンプリング) 75

artificial intelligence (人工知能) 3

ASO \triangleright alternating structural optimization

autoencoder (自己符号化器) 22, 91, 178

contractive –	25
deep –	92
denoising –	204
learning of –	22
marginalized denoising –	115
pre-training of –	93
recursive –	244
restricted Boltzmann machine as –	70
sparse –	24
stacked –	70
unfolding recursive –	246
variational –	112

average pooling (平均プーリング) 160

【 B 】

back-off smoothing (バックオフ平滑化) 229

backpropagation method (誤差逆伝播法) 14,
127

learning internal representation 16

backpropagation through time method 128, 212

bag-of-features model (bag-of-features モデル)
169

bag-of-words model (bag-of-words モデル)
226, 236

bag-of-visual-words model \triangleright bag-of-features
model

batch learning (バッチ学習) 13

batch normalization (バッチ正規化) 136

bi-directional recurrent neural network (双向
型回帰結合ニューラルネットワーク) 211

bigram (バイグラム) \triangleright *n*-gram

binary parse tree (二分構文木) 244

blocked Gibbs sampling (ブロック化ギブスサ
ンプリング) 55

BM \triangleright Boltzmann machine

BoF model \triangleright bag-of-features model

Boltzmann distribution (ボルツマン分布) 36

Boltzmann machine (ボルツマンマシン) 37
approximation technique of – 52

deep – 62

learning of – 40

learning equation of – 42

restricted – 57

BoW model \triangleright bag-of-words model

BPTT method \triangleright backpropagation through time
method

【C】

- CD method \Leftrightarrow contrastive divergence method
 CDBN \Leftrightarrow convolutional deep belief network
 central difference (中心差分) 147
 cerebral nerve system (脳神経系) 9, 176
 chain rule (連鎖律) 128
 CNN \Leftrightarrow convolutional neural network
 combinatorial explosion (組み合わせ爆発) 45
 competitive learning (競合学習) 15
 conditional independence (条件付き独立性) 59
 context (文脈) 229
 continuous bag-of-words model (連續 bag-of-words モデル) 235
 continuous skip-gram model (連續スキップグラム モデル) 237
 contractive autoencoder (縮小自己符号化器) 25
 contrastive divergence (コントラスティブ・ダイバージェンス) 103
 contrastive divergence method (コントラスティブ・ダイバージェンス法) 99
 loss function 100
 persistent – 109
 ConvNet \Leftrightarrow convolutional neural network
 convolution (畳み込み) 157
 convolution layer (畳み込み層) 157
 convolutional deep belief network (畳み込み深層信念ネットワーク) 181
 convolutional neural network (畳み込みニューラルネットワーク) 19, 156
 convolution layer 157
 difference from previous methods 170
 extension of – 174
 fully connected layer 157
 internal representation 175
 learning of – 163
 multitask learning 176
 pooling layer 160
 recognition performance 173
 corpus (コーパス) 228
 cost function (コスト関数) \Leftrightarrow loss function
 cross entropy (交差エントロピー) 163, 199

【D】

- DAE \Leftrightarrow deep autoencoder
 dark knowledge (暗黒知識) 138
 data parallelism (データ並列化) 130

- DBL method \Leftrightarrow detailed balance learning method
 DBM \Leftrightarrow deep Boltzmann machine
 DBN \Leftrightarrow deep belief network
 decoder (復号化器) 92
 deconvolutional network (逆畳み込みネットワーク) 175
 deep autoencoder (深層自己符号化器) 92
 deep belief network (深層信念ネットワーク) 72
 convolutional – 181
 inference of – 74
 pre-training of – 74
 deep Boltzmann machine (深層ボルツマンマシン) 62
 application of – 71
 learning of – 67
 pre-training of – 64
 deep hierarchical neural network (深層階層型ニューラルネットワーク) \Leftrightarrow deep neural network
 deep learning (深層学習) 3
 deep neural network (深層ニューラルネットワーク) 16
 acoustic model 206
 –hidden Markov model 201
 language model 218
 learning of – 17
 pre-training of – 202
 deep neural network-hidden Markov model (深層ニューラルネットワーク-隠れマルコフモデル) 201
 denoising autoencoder (雑音除去自己符号化器) 25, 115, 204
 dependency parsing (係り受け解析) 225
 detailed balance condition (詳細釣り合い条件) 105
 detailed balance learning method (detailed balance learning 法) 104
 dimension reduction (次元削減) 5, 8
 discriminative pre-training (識別的事前学習) 205
 DistBelief 130
 distillation (蒸留) 137
 distributed parallel computation (分散並列計算) 130

distributed semantic representation (分散的意味表現)	234	Fisher vector (フィッシャーベクトル)	170
distributional hypothesis (分布仮説)	233	fully connected layer (全結合層)	157
distributional memory (分布メモリ)	244	function word (機能語)	242
distributional semantic composition (分布的意味構築)	243	【 G 】	
distributional semantic representation (分布的意味表現)	234	Gabor filter (ガボールフィルタ)	171
distributional semantics (分布意味論)	229	Gaussian mixture model (混合正規分布)	192
divisive normalization (除算正規化)	165	Gaussian noise (ガウス雑音)	25
DNN \Leftrightarrow deep neural network		Gaussian-Bernoulli restricted Boltzmann machine (ガウス-ベルヌイ型制限ボルツマンマシン)	202
DNN-HMM \Leftrightarrow deep neural network-hidden Markov Model		general object recognition (一般物体認識)	153 , 166
Downpour SGD	131	previous methods for –	167
DropOut	139	generative model (生成モデル)	33
linear regression	141	Gibbs distribution (ギブス分布) \Leftrightarrow Boltzmann distribution	
dropout noise (欠落ノイズ) \Leftrightarrow masking noise		Gibbs sampler (ギブスサンプラー) \Leftrightarrow Gibbs sampling	
dynamic pooling (動的プーリング)	247	Gibbs sampling (ギブスサンプリング)	52
【 E 】			
EFH \Leftrightarrow exponential family harmonium		global feature (大域特徴)	167
ELM \Leftrightarrow extreme learning machine		GMM \Leftrightarrow Gaussian mixture model	
Elman network (エルマン・ネットワーク)	211	GPU	133
empirical distribution (経験分布)	43	gradient ascent method (勾配上昇法) \Leftrightarrow gradient descent method	
encoder (符号化器)	92	gradient descent method (勾配降下法)	45, 127
energy function (エネルギー関数)	36	grandmother cell (おばあさん細胞)	153
error correction learning (誤り訂正学習)	12	graphical model (グラフィカルモデル)	37
estimation (推定)	35	greedy layer-wise training (層ごとの貪欲学習)	
exchange Monte Carlo method (交換モンテカルロ法)	110	93 , 112	
exponential family harmonium (指数型ハーモニウム族)	96	grid search (グリッド探索)	146
Gibbs sampling	98	【 H 】	
extreme learning machine (極端学習機械)	116	harmonium (ハーモニウム) \Leftrightarrow restricted Boltzmann machine	
【 F 】			
FA \Leftrightarrow factor analysis		Heaviside function (ヘビサイド関数)	
factor analysis (因子分析)	8	\Leftrightarrow threshold function	
feature (特徴)	5	hidden layer (隠れ層)	11, 58
feature engineering (特徴工学)	7	hidden Markov model (隠れマルコフモデル)	
feature learning (特徴学習) \Leftrightarrow representation learning	representation	192	
feature point (特徴点)	65	deep neural network--	201
feed-forward neural network (フィードフォワード型ニューラルネットワーク) \Leftrightarrow hierarchical neural network		neural network--	195
hidden variable (隠れ変数)	40, 50, 94	hierarchical neural network (階層型ニューラルネットワーク)	8 , 127, 196

backpropagation method	14	language model (言語モデル)	192, 228		
deep –	☞ deep neural network	neural network –	229		
learning of –	12	large vocabulary continuous speech recognition (大語彙連続音声認識)	194		
– model	9	lasso	☞ L_1 regularization		
hierarchical softmax (階層型ソフトマックス)		latent representation (潜在表現)	☞ internal representation		
239		latent variable (潜在変数)	☞ hidden variable		
hinge loss (ヒンジ損失)	232	LCA	☞ local contrast normalization		
HMM	☞ hidden Markov model	learning rate (学習率)	12, 164		
HNN	☞ hierarchical neural network	tuning of –	144		
Hopfield network (ホップフィールド・ネットワーク)	39	LeNet	☞ convolutional neural network		
hourglass-type neural network (砂時計型ニューラルネットワーク)	22, 91	likelihood function (尤度関数)	41		
Huffman tree (ハフマン木)	239	limited-memory Broyden-Fletcher-Goldfarb-Shanno method	☞ L-BFGS method		
hyper parameter (超パラメータ)	146	linear separation (線形分離)	12		
【 I 】					
ICA	☞ independent component analysis	linearly separable (線形分離可能)	13		
identity function (恒等関数)	22	local contrast normalization (局所コントラスト正規化)	164		
ILSVRC	153	local feature (局所特徴)	167		
image recognition (画像認識)	153	local optimum (局所最適解)	15, 85		
independent component analysis (独立成分分析)	8	log Mel-filterbank feature vector (対数メルフィルタバンク特徴ベクトル)	203		
inference (推論)	35	log-bilinear form (対数双線形)	238		
InfiniBand	135	log-likelihood function (対数尤度関数)	41		
input layer (入力層)	11	long short-term memory method	214		
internal covariate shift (内部共変量シフト)	136	loss function (損失関数)	84, 127		
internal layer (中間層)	☞ hidden layer	low effective dimension	146		
internal representation (内部表現)	5, 16	L_p pooling (L_p プーリング)	160		
invariance (不变性)	7, 89, 160, 174, 183, 238	LSTM method	☞ long short-term memory method		
iterative parameter mixing method (反復的パラメータ混合法)	242	【 M 】			
【 K 】					
KL distance (KL 距離)	☞ Kullback-Leibler divergence	machine learning (機械学習)	4		
KL divergence (KL ダイバージェンス)	☞ Kullback-Leibler divergence	manifold learning (多様体学習)	8		
KL information (KL 情報量)	☞ Kullback-Leibler divergence	MAP estimation (MAP 推定)	☞ maximum a posteriori estimation		
Kullback-Leibler divergence (カルバック・ライbler・ダイバージェンス)	43, 97	marginalized denoising autoencoder (周辺化雑音除去自己符号化器)	115		
【 L 】					
L-BFGS method (L-BFGS 法)	131, 246	Markov chain Monte Carlo method (マルコフ連鎖モンテカルロ法)	52		
L_1 regularization (L_1 正則化)	25	Markov property (マルコフ性)	53		
【 M 】					
【 N 】					
masking noise (欠落雜音)	25	Markov random field (マルコフ確率場)	35		
max pooling (最大プーリング)	160				

- maximum a posteriori estimation (最大事後確率推定) 34
- maximum likelihood estimation (最尤推定) 41
- maximum likelihood estimator (最尤推定量) 41
- maximum mutual information criterion (最大相互情報量規準) 208
- MaxOut 143
- mean-field approximation (平均場近似) 55
- mean-field equation (平均場方程式) 57
- Mel-frequency cepstral coefficient (メル周波数ケプストラム係数) 193
- MFCC \Leftrightarrow mel-frequency cepstral coefficient
- mini batch method (ミニバッチ法) 129
- minimum probability flow method (最小確率流法) 108
- MLE \Leftrightarrow maximum likelihood estimation
- MMI criterion (MMI 規準) \Leftrightarrow maximum mutual information criterion
- model bias (モデルバイアス) \Leftrightarrow model error
- model error (モデル誤差) 51
- model parallelism (モデル並列化) 130
- molecular field approximation (分子場近似) \Leftrightarrow mean-field approximation
- moment matching (モーメント・マッチング) 42
- momentum (モメンタム) 164
- MPF method \Leftrightarrow minimum probability flow method
- MRF \Leftrightarrow Markov random field
- multi-layer neural network (多層ニューラルネットワーク) \Leftrightarrow hierarchical neural network
- multi-layer perceptron (多層パーセプトロン) \Leftrightarrow hierarchical neural network
- multi-stream learning (マルチストリーム学習) 217
- multitask learning (マルチタスク学習) 217, 233
- convolutional neural network 176
- mutually connected neural network (相互結合ニューラルネットワーク) 11
- MV-RNN \Leftrightarrow matrix vector recursive neural network
- natural language processing (自然言語処理) 225
- negative sampling (負例サンプリング) 241
- neocognitron (ネオコグニトロン) 19
- neural network (ニューラルネットワーク) 3
- convolutional – 19, 156
- deep – 16
- feed-forward – \Leftrightarrow hierarchical neural network
- –hidden Markov model 195
- hierarchical – 8
- history of – 9
- hourglass-type – 22
- multi-layer – \Leftrightarrow hierarchical neural network
- mutually connected – 11
- recurrent – 211
- time-delay – 197
- neural network language model (ニューラルネットワーク言語モデル) 229
- neural network–hidden Markov model (ニューラルネットワーク–隠れマルコフモデル) 195
- hybrid approach 198
- tandem approach 200
- neuron (ニューロン) 9
- n*-gram (*n* グラム) 194, 226, 228
- NLMM \Leftrightarrow neural network language model
- NN \Leftrightarrow neural network
- NN-HMM \Leftrightarrow neural network–hidden Markov Model
- no free lunch theorem (ノーフリーランチ定理) 6
- normalization constant (規格化定数) 37, 94
-
- [O]**
-
- objective function (目的関数) \Leftrightarrow loss function
- observable variable (観測変数) \Leftrightarrow visible variable
- one-versus-rest encoding (一対他符号化) \Leftrightarrow 1-of-*n* representation
- online learning (オンライン学習) 13
- output function (出力関数) \Leftrightarrow activation function
- output layer (出力層) 11
- over-fitting (過学習) 84, 139
-
- [N]**
-
- natural gradient method (自然勾配法) 88
-
- [P]**
-
- parallel tempering method (パラレル・テンパ

リング法)	\Rightarrow	exchange Monte Carlo method		
paraphrase detection (言い換え表現認識)		244		
part-of-speech tagging (品詞タグ付け)		225		
partition function (分配関数)	\Rightarrow	normalization constant		
PCA	\Rightarrow	principal component analysis		
perceptron (パーセプトロン)		9		
– learning	\Rightarrow	error correction learning		
– learning rule convergence theorem		13		
limitation of –		13		
multi-layer –	\Rightarrow	hierarchical neural network		
perceptron learning (パーセプトロン学習)				
– error correction learning				
persistent contrastive divergence method (継続的コントラスティブ・ダイバージェンス法)				
109				
phoneme (音素)		189		
phonemic context (音素文脈)		192		
plateau (プラトー)		85		
PoE	\Rightarrow	product of experts		
pooling (プーリング)		21, 160, 170		
average –		160		
dynamic –		247		
L_p –		160		
max –		160		
probabilistic max –		181		
pooling layer (プーリング層)		160		
post-training (事後学習)		227		
posterior distribution (事後分布)		34		
pre-training (事前学習)		23, 88, 112, 226		
– of deep belief network		74		
– of deep Boltzmann machine		64		
– of denoising autoencoder		204		
discriminative –		205		
pre-training of –		112		
– by probabilistic models		93		
stacked autoencoder		113		
principal component analysis (主成分分析)		8		
prior distribution (事前分布)		34		
probabilistic graphical model (確率的グラフィカルモデル)	\Rightarrow	graphical model		
probabilistic max pooling (確率的最大プーリング)		181		
product of experts		118		
pronunciation model (発音辞書モデル)		192		
pseudo negative instance (擬似負例)		242		
pyramid match kernel			170	
【R】				
random search (ランダム探索)			146	
RBM	\Rightarrow	restricted Boltzmann machine		
receptive field (受容野)			11, 159	
reconstruction error (再構成誤差)			92	
reconstruction TICA (再構成型 TICA)	\Rightarrow	reconstruction topographic independent component analysis		
reconstruction topographic independent component analysis (再構成型トポグラフィック独立成分分析)			182	
rectified linear unit			\Rightarrow ReLU	
recurrent neural network (回帰結合ニューラルネットワーク)				211
acoustic model				211
bi-directional –				211
Elman network				211
language model				218
long short-term memory method				214
recursive autoencoder (再帰自己符号化器)			244	
recursive neural network (再帰ニューラルネットワーク)				244
recursive autoencoder				244
unfolding recursive autoencoder				246
regret (リグレット)				145
regularization (正則化)				24
distillation				138
DropOut				142
L_1 –				25
sparse –				179
ReLU				143, 156
representation learning (表現学習)				7, 178, 226
restricted Boltzmann machine (制限ボルツマンマシン)				57, 94, 202
– as autoencoder				69
Bernoulli-Bernoulli –				57
conditional independence of –				59
Gaussian-Bernoulli –				202
learning of –				60
RNN	\Rightarrow	recurrent neural network, recursive neural network		
robustness (ロバスト性)				7

【S】

saddle point (鞍点)	85	topographic independent component analysis （トポグラフィック独立成分分析）	182
salt-and-pepper noise (ごま塩雜音)	25	reconstruction –	182
Sandblaster L-BFGS	131	topographic mapping (トポグラフィック・マッピング)	183
scale invariant feature transform	☞ SIFT	transfer learning (転移学習)	☞ multitask learning
semantic representation (意味表現)	232, 233	trigram (トライグラム)	☞ n-gram
sequence discriminative training (系列識別學習)	207		
SGD	☞ stochastic gradient descent method		
shared weight (共有重み)	159		
SIFT	168		
sigmoid belief (シグモイド信念)	39		
sigmoid function (シグモイド関数)	14		
simple perceptron (単純バーセプトロン)	12		
smoothing (平滑化)	229		
softmax function (ソフトマックス関数)	157		
– output layer	198		
sparse autoencoder (スパース自己符号化器)	24, 179		
sparsereness (スパース性)	7, 8		
speaker adaptation (話者適応)	200		
speech recognition (音声認識)	190		
stacked autoencoder (積層自己符号化器)	23, 70, 113		
statistical machine learning (統計的機械学習)	33		
steepest gradient descent method (最急勾配降下法)	☞ gradient descent method		
stochastic gradient descent method (確率的勾配降下法)	14, 128		
stop word (不要語)	242		
subtractive normalization (減算正規化)	164		
supervised learning (教師あり学習)	12		
synapse (シナプス)	9		

【T】

TDNN	☞ time-delay neural network
temperature parameter (温度パラメータ)	110
threshold function (しきい関数)	10
TICA	☞ topographic independent component analysis
tied weight	☞ shared weight
time-delay neural network (時間遅れニューラルネットワーク)	197

topographic independent component analysis （トポグラフィック独立成分分析）	182
reconstruction –	182

topographic mapping (トポグラフィック・マッピング)	183
--------------------------------------	-----

transfer learning (転移学習)	☞ multitask learning
--------------------------	----------------------

trigram (トライグラム)	☞ n-gram
------------------	----------

【U】

ugly duckling theorem (醜いアヒルの子の定理)	6
------------------------------------	---

undirected graph (無向グラフ)	36
--------------------------	----

unfolding recursive autoencoder (展開再帰自己符号化器)	246
--	-----

unigram (ユニグラム)	☞ n-gram
-----------------	----------

unsupervised learning (教師なし学習)	12
--------------------------------	----

【V】

vanishing gradient problem (勾配消失問題)	18, 214
-------------------------------------	---------

variational autoencoder (変分自己符号化器)	112
------------------------------------	-----

variational Bayes method (変分ベイズ法)	☞ mean-field approximation
-----------------------------------	----------------------------

visible layer (可視層)	58
---------------------	----

visible variable (可視変数)	40, 94
-------------------------	--------

visual word (ビジュアルワード)	169
------------------------	-----

【W】

weight (重み)	10
-------------	----

weight decay (重み減衰)	164
---------------------	-----

weight sharing (重み共有)	159
-----------------------	-----

whitening (白色化)	165
-----------------	-----

wine-glass-type neural network (ワイングラス型ニューラルネットワーク)	☞ hourglass-type neural network
---	---------------------------------

word error rate (単語エラー率)	207
--------------------------	-----

word2vec	235
----------	-----

【Z】

zero-frequency problem (ゼロ頻度問題)	194,
---------------------------------	------

228

執筆者紹介

第 1 章

麻生 英樹 (あそう ひでき)

産業技術総合研究所人工知能研究センター，副研究センター長
機械学習とその応用

第 2 章

安田 宗樹 (やすだ むねき)

山形大学理工学研究科，准教授
機械学習，情報統計力学，画像処理

第 3 章

前田 新一 (まえだ しんいち)

京都大学大学院情報学研究科，助教
機械学習，強化学習，統計的画像処理

第 4 章

岡野原 大輔 (おかのはら だいすけ)

株式会社 Preferred Infrastructure, 株式会社 Preferred Networks, 取締役副社長
機械学習，大規模データ解析，自然言語処理，データ構造

第 5 章

岡谷 貴之 (おかたに たかゆき)

東北大学情報科学研究科，教授
コンピュータビジョン

第 6 章

久保 陽太郎 (くぼ ようたろう)

Amazon, Speech Scientist
音声認識

第 7 章

ボレガラ ダヌシカ

University of Liverpool, Department of Computer Science, Associate Professor
自然言語処理，データマイニング，機械学習

編集

神島 敏弘 (かみしま としひろ)

産業技術総合研究所，主任研究員
データマイニング，機械学習，推薦システム

監修：人工知能学会 代表

松尾 豊 (まつお ゆたか)

東京大学大学院工学系研究科、准教授、前人工知能学会編集委員長
人工知能、ウェブ工学、ビッグデータ分析

松原 仁 (まつばら ひとし)

公立はこだて未来大学システム情報科学部、教授、人工知能学会会長
人工知能

深層学習 –Deep Learning–

© 2015 Toshihiro Kamishima, Hideki Asoh,
Muneki Yasuda, Shin-ichi Maeda, Daisuke
Okanohara, Takayuki Okatani, Yotaro Kubo,
Danushka Bollegala

Printed in Japan

2015年10月31日 初版第1刷発行

監修者 人工知能学会

編 者 神 嶽 敏 弘

共著者 麻 生 英 樹

安 田 宗 樹

前 田 新 一

岡 野 原 大 輔

岡 谷 貴 之

久 保 陽 太 郎

ボレガラ ダヌシカ

発行者 小 山 透

発行所 株式会社 近代科学社

〒 162-0843 東京都新宿区市谷田町 2-7-15

電話 03-3260-6161 振替 00160-5-7625

<http://www.kindaikagaku.co.jp>

藤原印刷 ISBN978-4-7649-0487-3

定価はカバーに表示しております。

ISBN978-4-7649-0487-3
C3004 ￥3500E



9784764904873

定価(本体 3,500円+税)



1923004035009

Deep Learning

深層學習



近代科学社