

深 層 学 習

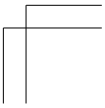
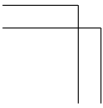
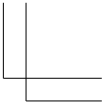
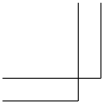
Deep Learning

編集

人工知能学会，神嶋 敏弘

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



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。というのも、人間の脳は、多段につみかさ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ひと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、各章ともに分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。そういった方には、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1章から4章までで構成され、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付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から始まり、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、深層学習におけるひと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トティブ・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付けからはじまり、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ハイパー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関して解説されている。

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。両者の位置付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ひと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。

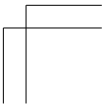
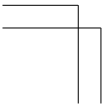
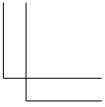
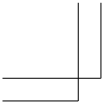
第7章は、ボレガラ・ダヌシカによる自然言語処理分野における解説であ

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始まり、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なのが、深層学習はまだ始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々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に感謝したい：《お世話になった方の名前》(敬称略)

松尾 豊



目次

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 (麻生 英樹)	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	11
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	17
1.5.2 深層ニューラルネットワークの学習	17
1.6 畳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	22
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	26

参考文献	28
第2章 深層ボルツマンマシン (安田 宗樹)	33
2.1 はじめに	33
2.2 統計的機械学習の考え方 — データ生成モデルの再現	34
2.3 マルコフ確率場とボルツマンマシン	37
2.3.1 マルコフ確率場	38
2.3.2 ボルツマンマシン	39
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	41
2.3.4 ボルツマンマシンの学習へ	42
2.4 可視変数のみのボルツマンマシン学習	42
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	45
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	47
2.5 隠れ変数ありのボルツマンマシン学習	48
2.5.1 隠れ変数ありの場合の学習について	51
2.5.2 隠れ変数を導入する意味	52
2.6 ボルツマンマシン上での近似手法	54
2.6.1 ギブスサンプリング	55
2.6.2 平均場近似	58
2.7 制限ボルツマンマシン	60
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	61
2.7.2 制限ボルツマンマシンの学習	62
2.8 深層ボルツマンマシン	64
2.8.1 深層ボルツマンマシンの事前学習	66
2.8.2 事前学習後の最尤推定法に基づく学習	70
2.8.3 自己符号化器としての制限ボルツマンマシン	72
2.8.4 深層ボルツマンマシンの利用法	73
2.9 深層信念ネットワーク	74
2.9.1 深層信念ネットワークに対する事前学習と推論	76
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	77
2.10 おわりに	82
参考文献	82

第 3 章	事前学習とその周辺（前田 新一）	85
3.1	はじめに	85
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 . . .	86
3.2.1	学習を難しくする要因	86
3.2.2	既存の解決法	88
3.2.3	新たな解決法	90
3.3	自己符号化器による内部表現の学習	93
3.3.1	自己符号化器とその損失関数の定義	93
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 . .	95
3.4	確率的なモデルを用いた事前学習	96
3.4.1	制限ボルツマンマシン	96
3.4.2	指数型ハーモニウム族	98
3.4.3	指数型ハーモニウム族のコントラストティブ・ダイバー ジェンス法による学習	102
3.4.4	コントラストティブ・ダイバージェンス法が最適化して いる損失関数	103
3.4.5	コントラストティブ・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	110
3.4.6	コントラストティブ・ダイバージェンス法から派生した 学習則	112
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	113
3.5	確定的なモデルを用いた事前学習	115
3.5.1	教師なし学習による確定的なモデルの学習	115
3.5.2	教師あり学習による確定的なモデルの学習	119
3.6	Product of Experts の学習法としてのコントラストティブ・ダイ バージェンス法	120
3.7	おわりに	121
	参考文献	122
第 4 章	大規模深層学習の実現技術（岡野原 大輔）	127
4.1	はじめに	127
4.2	深層学習の最適化	129
4.2.1	深層学習の基本計算	129
4.2.2	確率的勾配降下法	131
4.3	高速化手法	131

4.3.1	分散並列処理：DistBelief	132
4.3.2	GPU を用いた大規模ニューラルネットの実現	135
4.3.3	InfiniBand の利用	137
4.3.4	学習収束の高速化	138
4.4	過学習制御：DropOut	141
4.5	活性化関数	145
4.5.1	ReLU	145
4.5.2	MaxOut	145
4.6	学習率の調整	147
4.6.1	AdaGrad	147
4.6.2	Adam	147
4.6.3	超パラメータの最適化	148
4.7	実装技術	149
4.7.1	実装の正しさのチェック	149
4.8	おわりに	150
	参考文献	150

第 II 部 応用編 153

第 5 章	画像認識のための深層学習（岡谷 貴之）	155
5.1	はじめに	155
5.1.1	畳み込みニューラルネットワークの再発見	156
5.1.2	その後の研究	156
5.2	畳み込みニューラルネットワーク	158
5.2.1	基本構造	158
5.2.2	畳み込み層	159
5.2.3	プーリング層	161
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	163
5.2.5	学習	165
5.2.6	コントラスト調整とデータの正規化	166
5.3	畳み込みニューラルネットワークのはたらき	168
5.3.1	一般物体認識の難しさ	168
5.3.2	一般物体認識の従来法	169
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 . . .	172

5.3.4	ネットワークの構造と認識性能	175
5.3.5	畳み込みニューラルネットワークの拡張の試み	176
5.4	畳み込みニューラルネットワークの内部表現	177
5.4.1	可視化	177
5.4.2	脳神経系との関係	178
5.4.3	転移学習	178
5.5	画像特徴の教師なし学習	180
5.5.1	単層自己符号化器による局所特徴の学習	180
5.5.2	多層ネットワークによる特徴学習	183
5.6	おわりに	186
	参考文献	187
第 6 章	音声認識のための深層学習 (久保 陽太郎)	193
6.1	はじめに	193
6.2	音声認識	195
6.2.1	音声認識に用いられるモデル	195
6.2.2	大語彙連続音声認識システムの構成	198
6.3	音声認識におけるニューラルネットワーク	199
6.3.1	時間遅れニューラルネットワーク	201
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	202
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	204
6.4	音響モデルにおける深層学習: 事前学習	206
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	206
6.4.2	雑音除去自己符号化器による事前学習	208
6.4.3	識別的な事前学習	209
6.5	音響モデルにおける深層学習: 学習とモデルの進展	211
6.5.1	系列識別学習	211
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	215
6.5.3	Long Short-Term Memory 法	218
6.5.4	マルチストリーム・マルチタスク学習	221
6.6	言語モデルにおける深層学習	222
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	223

6.7	おわりに	225
	参考文献	226
第 7 章	自然言語処理のための深層学習 (ボレガラ ダヌシカ)	229
7.1	はじめに	229
7.2	深層学習と言語モデル	232
7.2.1	ニューラルネットワーク言語モデル	233
7.2.2	その他の言語モデル	235
7.3	単語の意味表現学習	237
7.3.1	ボトムアップ的な意味表現構築手法	237
7.3.2	トップダウン的な意味表現予測手法	238
7.3.3	階層型ソフトマックスによる計算	243
7.3.4	意味表現学習手法のその他の話題	244
7.4	深層学習と意味構築	247
7.4.1	言い換え表現認識への応用	248
7.5	おわりに	252
	参考文献	253

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 (大文字ローマン体)	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
\mathcal{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)
$\theta, \boldsymbol{\theta}, \boldsymbol{\Theta}$	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 (要素が全て 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
X^T	行列 X の転置
X^{-1}	行列 X の逆行列
$X \circ Y$	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
$\text{diag}(d)$	ベクトル 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。

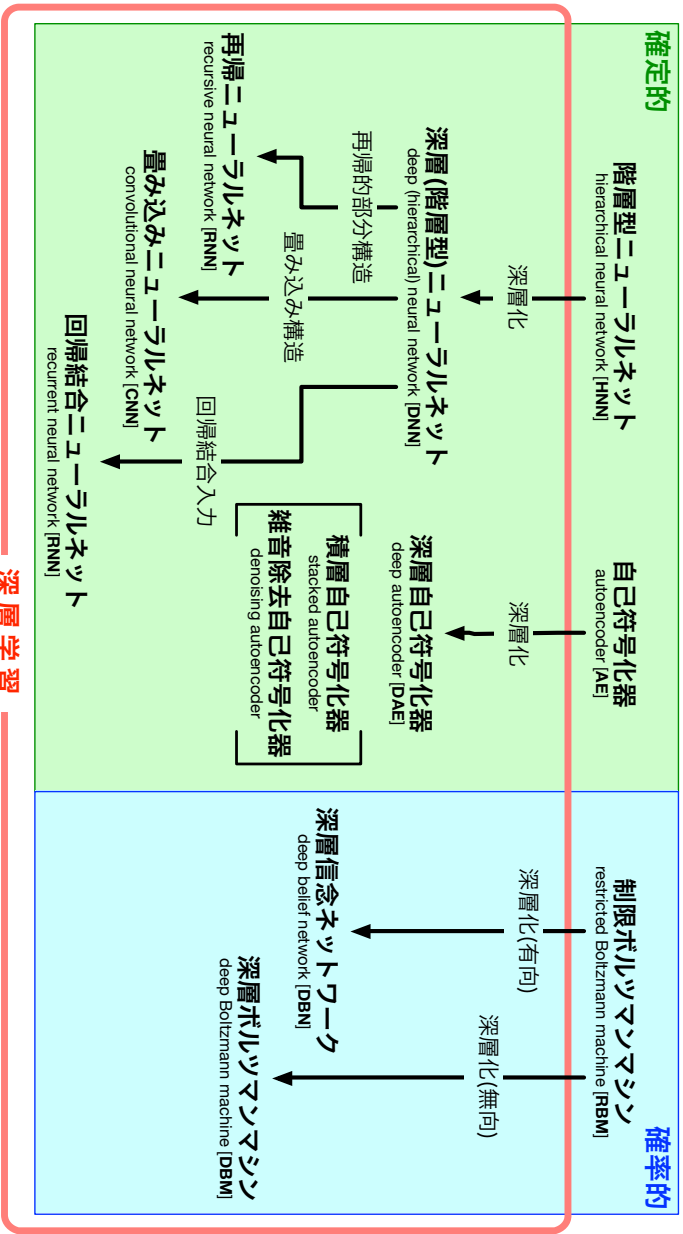
期待値 $E_{p(X)}[f(X)]$ は、分布 $p(X)$ についての次の期待値を表す：

$$\begin{aligned} \sum_{x \in \text{Dom}(X)} f(x)p(x) & \quad \text{— } X \text{ が離散の場合} \\ \int_{x \in \text{Dom}(X)} f(x)p(x)dx & \quad \text{— } X \text{ が連続の場合} \end{aligned}$$

なお、 $p(X)$ を省略した場合は、関数 f の全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：

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



深層学習手法の全体像

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付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』に基づいて、2種類の観点から各種モデルの分類を試みる。図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分かる観測変数と、値が分からない隠れ変数があり、そ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

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。

パーセプトロン【1.4.1 節】

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、多層パーセプトロンとも呼ばれ、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指す。この狭義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根拠と、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、中間ノード数の増加より、深層化の方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 節】

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 節】

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。そのために、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

もう一方の自己符号化器は、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 節, 1.7 節】

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、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 節】

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 節】

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 節, 6.4.2 節】

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 節】

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は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層とみなすと、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ト・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。深層ボルツマンマシンと同様の技術が適用できることから、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)	135, 165, 233
Adagrad	147
Adam	147
ASO	⇨ 交互構造最適化
backpropagation through time 法	130, 216
bag-of-features モデル (bag-of-features model)	172
bag-of-words モデル (bag-of-words model)	230, 240
bag-of-visual-words モデル	⇨ bag-of-features モデル
BM	⇨ ボルツマンマシン
BoF モデル	⇨ bag-of-features モデル
BoW モデル	⇨ bag-of-words モデル
BPTT 法	⇨ backpropagation through time 法
CDBN	⇨ 畳み込み深層信念ネットワーク
CD 法	⇨ コントラストティブ・ダイバージェンス法
CNN	⇨ 畳み込みニューラルネットワーク
ConvNet	⇨ 畳み込みニューラルネットワーク
DAE	⇨ 深層自己符号化器
DBL 法	⇨ detailed balance learning 法
DBM	⇨ 深層ボルツマンマシン
DBN	⇨ 深層信念ネットワーク
detailed balance learning 法 (detailed balance learning method)	106
DistBelief	132
DNN	⇨ 深層ニューラルネットワーク
DNN-HMM	⇨ 深層ニューラルネットワーク-隠れマルコフモデル
Downpour SGD	133
DropOut	141

線形回帰	144
EFH	⇨ 指数型ハーモニウム族
ELM	⇨ 極端学習機械
FA	⇨ 因子分析
GMM	⇨ 混合正規分布
GPU	135
HMM	⇨ 隠れマルコフモデル
HNN	⇨ 階層型ニューラルネットワーク
ICA	⇨ 独立成分分析
ILSVRC	155
InfiniBand	137
KL 距離 (KL distance)	⇨ カルバック-ライブラー・ダイバージェンス
KL 情報量 (KL information)	⇨ カルバック-ライブラー・ダイバージェンス
KL ダイバージェンス (KL divergence)	⇨ カルバック-ライブラー・ダイバージェンス
L-BFGS 法 (L-BFGS method)	134, 250
L_1 正則化 (L_1 regularization)	25
lasso	⇨ L_1 正則化
LCA	⇨ 局所コントラスト正規化
LeNet	⇨ 畳み込みニューラルネットワーク
limited-memory Broyden-Fletcher-Goldfarb-Shanno 法	⇨ L-BFGS 法
long short-term memory 法	218
low effective dimension	148
L_p プーリング (L_p pooling)	162
LSTM 法	⇨ long short-term memory 法
MAP 推定 (MAP estimation)	⇨ 最大事後確率推定
MaxOut	145
MFCC	⇨ メル周波数ケプストラム係数
MLE	⇨ 最尤推定
MMI 規準 (MMI criterion)	⇨ 最大相互情報量

規準	
MPF 法	⇒ 最小確率流法
MRF	⇒ マルコフ確率場
MV-RNN	⇒ matrix vector recursive neural network
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
n グラム (n -gram)	198, 230, 232
PCA	⇒ 主成分分析
PoE	⇒ product of experts
product of experts	120
pyramid match kernel	172
RBM	⇒ 制限ボルツマンマシン
rectified linear unit	⇒ ReLU
ReLU	145, 159
RNN	⇒ 回帰結合ニューラルネットワーク, 回帰ニューラルネットワーク
Sandblaster L-BFGS	133
scale invariant feature transform	⇒ SIFT
SGD	⇒ 確率的勾配降下法
SIFT	170
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	239

【あ】

誤り訂正学習 (error correction learning)	12
暗黒知識 (dark knowledge)	141
鞍点 (saddle point)	87
言い換え表現認識 (paraphrase detection)	248
一対他符号化 (one-versus-rest encoding) ⇒ 1-of- n 表現	
一般物体認識 (general object recognition)	155, 168
～の従来法	169
意味表現 (semantic representation)	236, 237
因子分析 (factor analysis)	8
エネルギー関数 (energy function)	38
エルマン・ネットワーク (Elman network)	215
おばあさん細胞 (grandmother cell)	156
重み (weight)	10
重み共有 (weight sharing)	161
重み減衰 (weight decay)	166

音響モデル (acoustic model)	195
音声認識 (speech recognition)	195
音素 (phoneme)	194
音素文脈 (phonemic context)	196
温度パラメータ (temperature parameter)	113
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)	215
long short-term memory 法	218
エルマン・ネットワーク	215
音響モデル	215
言語モデル	223
双方向型～	215
階層型ソフトマックス (hierarchical softmax)	243
階層型ニューラルネットワーク (hierarchical neural network)	8, 129, 200
～の学習	11
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)	206
ガウス雑音 (Gaussian noise)	25
過学習 (over-fitting)	86, 141
係り受け解析 (dependency parsing)	229
学習率 (learning rate)	13, 166
～の調整	147
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, 131
確率的最大プーリング (probabilistic max pooling)	183
隠れ層 (hidden layer)	11, 60
隠れ変数 (hidden variable)	42, 52, 96
隠れマルコフモデル (hidden Markov model)	195
深層ニューラルネットワーク～	206
ニューラルネットワーク～	199
可視層 (visible layer)	60
可視変数 (visible variable)	42, 96
画像認識 (image recognition)	155

活性化関数 (activation function)	10, 145, 161	交互構造最適化 (alternating structural optimization)	231
ReLU～	145	交互最適化 (alternating optimization)	242
恒等～	22	交差エントロピー (cross entropy)	165, 203
しきい～	10	恒等関数 (identity function)	22
シグモイド～	14	勾配降下法 (gradient descent method)	47, 129
ソフトマックス～	159	勾配消失問題 (vanishing gradient problem)	18, 218
ガボールフィルタ (Gabor filter)	173	勾配上昇法 (gradient ascent method) ⇨ 勾配降下法	
カルバック-ライブラー・ダイバージェンス (Kullback-Leibler divergence)	45, 99	コーパス (corpus)	232
頑健性 (robustness) ⇨ ロバスト性		誤差逆伝播法 (backpropagation method)	14, 130
観測変数 (observable variable) ⇨ 可視変数		内部表現学習	17
機械学習 (machine learning)	4	コスト関数 (cost function) ⇨ 損失関数	
規格化定数 (normalization constant)	39, 96	ごま塩雑音 (salt-and-pepper noise)	25
擬似負例 (pseudo negative instance)	246	混合正規分布 (Gaussian mixture model)	197
機能語 (function word)	246	コントラストティブ・ダイバージェンス (contrastive divergence)	105
ギブスサンプラー (Gibbs sampler) ⇨ ギブスサンプリング		コントラストティブ・ダイバージェンス法 (contrastive divergence method)	102
ギブスサンプリング (Gibbs sampling)	55	継続的～	112
ギブス分布 (Gibbs distribution) ⇨ ボルツマン分布		損失関数	103
逆畳み込みネットワーク (deconvolutional network)	177		
競合学習 (competitive learning)	15	【さ】	
教師あり学習 (supervised learning)	12	再帰自己符号化器 (recursive autoencoder)	248
教師なし学習 (unsupervised learning)	12	再帰ニューラルネットワーク (recursive neural network)	
共有重み (shared weight)	161	再帰自己符号化器	248
局所コントラスト正規化 (local contrast normalization)	166	展開再帰自己符号化器	250
局所最適解 (local optimum)	15, 87	最急勾配降下法 (steepest gradient descent method) ⇨ 勾配降下法	
局所特徴 (local feature)	169	再構成型 TICA (reconstruction TICA) ⇨ 再構成型トポグラフィック独立成分分析	
極端学習機械 (extreme learning machine)	118	再構成型トポグラフィック独立成分分析 (reconstruction topographic independent component analysis)	184
組み合わせ爆発 (combinatorial explosion)	47	再構成誤差 (reconstruction error)	94
グラフィカルモデル (graphical model)	39	最小確率流法 (minimum probability flow method)	110
グリッド探索 (grid search)	148	最大事後確率推定 (maximum a posteriori estimation)	37
経験分布 (empirical distribution)	45	最大相互情報量規準 (maximum mutual information criterion)	212
継続的コントラストティブ・ダイバージェンス法 (persistent contrastive divergence method)	112	最大プーリング (max pooling)	162
系列識別学習 (sequence discriminative training)	211	最尤推定 (maximum likelihood estimation)	43
欠落雑音 (masking noise)	25		
言語モデル (language model)	195, 232		
ニューラネットワーク～	233		
減算正規化 (subtractive normalization)	166		
交換モンテカルロ法 (exchange Monte Carlo method)	113		

最尤推定量 (maximum likelihood estimator)	43	出力層 (output layer)	11
雑音除去自己符号化器 (denoising autoencoder)	25, 117, 208	受容野 (receptive field)	11 , 161
時間遅れニューラルネットワーク (time-delay neural network)	201	条件付き独立性 (conditional independence)	61
しきい関数 (threshold function)	10	詳細釣り合い条件 (detailed balance condition)	107
識別的事前学習 (discriminative pre-training)	209	蒸留 (distillation)	140
シグモイド関数 (sigmoid function)	14	除算正規化 (divisive normalization)	167
シグモイド信念 (sigmoid belief)	41	人工知能 (artificial intelligence)	3
次元削減 (dimension reduction)	5 , 8	深層階層型ニューラルネットワーク (deep hierarchical neural network)	☞ 深層ニューラルネットワーク
事後学習 (post-training)	231	深層学習 (deep learning)	3
自己符号化器 (autoencoder)	22, 93 , 180	深層自己符号化器 (deep autoencoder)	94
～の学習	22	深層信念ネットワーク (deep belief network)	74
再帰～	248	～の事前学習	76
雑音除去～	208	～の推論	76
～の事前学習	95	畳み込み～	183
周辺化雑音除去～	118	深層ニューラルネットワーク (deep neural network)	16
縮小～	26	音響モデル	211
深層～	94	～の学習	17
スパース～	24	～隠れマルコフモデル	206
～としての制限ボルツマンマシン	72	言語モデル	222
積層～	72	～の事前学習	206
展開再帰～	250	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)	206
変分～	114	深層ボルツマンマシン (deep Boltzmann machine)	64
事後分布 (posterior distribution)	36	～の学習	70
指数型ハロモニウム族 (exponential family harmonium)	98	～の事前学習	66
ギブスサンプリング	100	～の利用法	73
事前学習 (pre-training)	22, 90 , 115, 230	推定 (estimation)	37
確率的モデルによる～	96	推論 (inference)	37
雑音除去自己符号化器の～	208	砂時計型ニューラルネットワーク (hourglass-type neural network)	22 , 93
識別的～	209	スパース自己符号化器 (sparse autoencoder)	24, 181
自己符号化器の～	115	スパース性 (sparseness)	8
深層信念ネットワークの～	76	正規化定数 (normalization constant)	☞ 規格化定数
深層ボルツマンマシンの～	66	制限ボルツマンマシン (restricted Boltzmann machine)	60 , 96, 206
積層自己符号化器	115	ガウス-ベルヌーイ型～	206
自然言語処理 (natural language processing)	229		
自然勾配法 (natural gradient method)	90		
事前分布 (prior distribution)	36		
シナプス (synapse)	9		
周辺化雑音除去自己符号化器 (marginalized denoising autoencoder)	118		
縮小自己符号化器 (contractive autoencoder)	26		
主成分分析 (principal component analysis)	8		
出力関数 (output function)	☞ 活性化関数		

～の学習	62	畳み込みニューラルネットワーク (convolu-	
自己符号化器としての～	72	tional neural network)	19,
～の条件付き独立性	61	158	
ベルヌーイ-ベルヌーイ型～	60	～の学習	165
生成モデル (generative model)	35	～の拡張	176
正則化 (regularization)	25	従来法との比較	172
DropOut	144	全結合層	159
L_1 ～	25	畳み込み層	159
蒸留	140	内部表現	177
スパース～	181	認識性能	175
積層自己符号化器 (stacked autoencoder)	22, 72,	プーリング層	161
115		マルチタスク学習	178
ゼロ頻度問題 (zero-frequency problem)	198, 233	多様体学習 (manifold learning)	8
線形分離 (linear separation)	12	単語エラー率 (word error rate)	211
線形分離可能 (linearly separable)	13	単語集合モデル (bag-of-words model) ⇨ bag-of-	
全結合層 (fully connected layer)	159	words モデル	
潜在表現 (latent representation) ⇨ 内部表現		単語袋詰めモデル (bag-of-words model) ⇨ bag-	
潜在変数 (latent variable) ⇨ 隠れ変数		of-words モデル	
相互結合ニューラルネットワーク (mutually		単純パーセプトロン (simple perceptron)	12
connected neural network)	10	中間層 (internal layer) ⇨ 隠れ層	
層ごとの貪欲学習 (greedy layer-wise training)		中心差分 (central difference)	150
95, 115		超パラメータ (hyper parameter)	148
双方向型回帰結合ニューラルネットワーク (bi-		通時的誤差逆伝播法 ⇨ backpropagation through	
directional recurrent neural network)	215	time 法	
疎性 (sparseness) ⇨ スパース性		データ並列化 (data parallelism)	132
ソフトマックス関数 (softmax function)	159	転移学習 (transfer learning) ⇨ マルチタスク	
～出力層	203	学習	
損失関数 (loss function)	86, 129	展開再帰自己符号化器 (unfolding recursive au-	
【た】		toencoder)	250
大域特徴 (global feature)	169	伝承サンプリング (ancestral sampling)	76
大語彙連続音声認識 (large vocabulary continu-		統計的機械学習 (statistical machine learning)	35
ous speech recognition)	198	動的プーリング (dynamic pooling)	251
対数双線形 (log-bilinear form)	242	特徴 (feature)	5
対数メルフィルタバンク特徴ベクトル (log mel-		特徴学習 (feature learning) ⇨ 表現学習	
filterbank feature vector)	207	特徴工学 (feature engineering)	7
対数尤度関数 (log-likelihood function)	44	特徴点 (feature point)	67
多層ニューラルネットワーク (multi-layer neural		独立成分分析 (independent component analysis)	
network) ⇨ 階層型ニューラルネットワーク		8	
多層パーセプトロン (multi-layer perceptron)		トポグラフィック・マッピング (topographic	
⇨ 階層型ニューラルネットワーク		mapping)	185
畳み込み (convolution)	159	トポグラフィック独立成分分析 (topographic	
畳み込み深層信念ネットワーク (convolutional		independent component analysis)	184
deep belief network)	183	再構成型～	184
畳み込み層 (convolution layer)	159	トライグラム (trigram) ⇨ n グラム	

【な】

内部共変量シフト (internal covariate shift)	138
内部表現 (internal representation)	5, 17
二分構文木 (binary parse tree)	248
ニューラルネットワーク (neural network)	3
回帰結合～	215
階層型～	8
～隠れマルコフモデル	199
時間遅れ～	201
深層～	16
砂時計型～	22
相互結合～	10
多層～ ㊦ 階層型ニューラルネットワーク	
畳み込み～	19, 158
フィードフォワード型～ ㊦ 階層型ニューラルネットワーク	
～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)	199
タンデム方式	204
ハイブリッド方式	202
ニューラルネットワーク言語モデル (neural network language model)	233
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)	㊦ 雑音
脳神経系 (cerebral nerve system)	9, 178
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 ㊦ 誤り訂正学習	
～学習則の収束定理	13
～の限界	14
多層～ ㊦ 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) ㊦ 誤り訂正学習	
ハーモニウム (harmonium) ㊦ 制限ボルツマンマシン	
バイグラム (bigram) ㊦ n グラム	
白色化 (whitening)	167
発音辞書モデル (pronunciation model)	195
バックオフ平滑化 (back-off smoothing)	233
バッチ学習 (batch learning)	13

バッチ正規化 (batch normalization)	138
ハフマン木 (Huffman tree)	243
パラレル・テンパリング法 (parallel tempering method) ㊦ 交換モンテカルロ法	
反復のパラメータ混合法 (iterative parameter mixing method)	246
ビジュアルワード (visual word)	170
表現学習 (representation learning)	7, 180, 230
ヒンジ損失 (hinge loss)	236
品詞タグ付け (part-of-speech tagging)	229
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) ㊦ 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)	172
プーリング (pooling)	21, 161, 172
L_p ～	162
確率的最大～	183
最大～	162
動的～	251
平均～	162
プーリング層 (pooling layer)	161
復号化器 (decoder)	94
符号化器 (encoder)	94
不変性 (invariance)	7, 91, 161, 176, 185, 242
不用語 (stop word)	246
プラトー (plateau)	87
負例サンプリング (negative sampling)	245
ブロック化ギブスサンプリング (blocked Gibbs sampling)	57
分散の意味表現 (distributed semantic representation)	238
分散並列計算 (distributed parallel computation)	132
分子場近似 (molecular field approximation) ㊦ 平均場近似	
分配関数 (partition function) ㊦ 規格化定数	
分布意味論 (distributional semantics)	233
分布仮説 (distributional hypothesis)	237
分布の意味構築 (distributional semantic composition)	247
分布の意味表現 (distributional semantic representation)	238
分布メモリ (distributional memory)	248
文脈 (context)	233
平滑化 (smoothing)	233

平均場近似 (mean-field approximation)	58
平均場方程式 (mean-field equation)	59
平均プーリング (average pooling)	162
ヘビサイド関数 (Heaviside function) ⇨ しきい関数	
変分自己符号化器 (variational autoencoder)	114
変分ベイズ法 (variational Bayes method) ⇨ 平均場近似	
ホップフィールド・ネットワーク (Hopfield network)	41
ボルツマン分布 (Boltzmann distribution)	39
ボルツマンマシン (Boltzmann machine)	39
～の学習	42
～の学習方程式	44
～の近似手法	54
深層～	64
制限～	60

【ま】

マルコフ確率場 (Markov random field)	38
マルコフ性 (Markov property)	55
マルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo method)	55
マルチストリーム学習 (multi-stream learning)	221
マルチタスク学習 (multitask learning) 222, 237	
畳み込みニューラルネットワーク	178
醜いアヒルの子の定理 (ugly duckling theorem)	6
ミニバッチ法 (mini batch method)	131
無向グラフ (undirected graph)	38
メル周波数ケプストラム係数 (mel-frequency cepstral coefficient)	197
モーメント・マッチング (moment matching)	45
目的関数 (objective function) ⇨ 損失関数	
モデル誤差 (model error)	53
モデルバイアス (model bias) ⇨ モデル誤差	
モデル並列化 (model parallelism)	132
モメンタム (momentum)	166

【や】

尤度関数 (likelihood function)	43
ユニグラム (unigram) ⇨ n グラム	

【ら】

ランダム探索 (random search)	148
------------------------	-----

リグレット (regret)	147
類推問題 (analogical reasoning problem)	246
連鎖律 (chain rule)	130
連続 bag-of-words モデル (continuous bag-of-words model)	239
連続スキップグラムモデル (continuous skip-gram model)	241
ロス関数 (loss function) ⇨ 損失関数	
ロバスト性 (robustness)	7

【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)	204

英文索引

【 Symbols / Numbers 】

1-of- n representation (1-of- n 表現) 135, 165, 233

【 A 】

acoustic model (音響モデル) 195
activation function (活性化関数) 10, 145, 161
 identity – 22
 ReLU – 145
 sigmoid – 14
 softmax – 159
 threshold – 10
Adagrad 147
Adam 147
alternating optimization (交互最適化) 242
alternating structural optimization (交互構造最適化) 231
analogical reasoning problem (類推問題) 246
ancestral sampling (伝承サンプリング) 76
artificial intelligence (人工知能) 3
ASO ⇔ alternating structural optimization
autoencoder (自己符号化器) 22, 93, 180
 contractive – 26
 deep – 94
 denoising – 208
 learning of – 22
 marginalized denoising – 118
 pre-training of – 95
 recursive – 248
 restricted Boltzmann machine as – 72
 sparse – 24
 stacked – 72
 unfolding recursive – 250
 variational – 114

average pooling (平均プーリング) 162

【 B 】

back-off smoothing (バックオフ平滑化) 233
backpropagation method (誤差逆伝播法) 14, 130
 learning internal representation 17
backpropagation through time method 130, 216
bag-of-features model (bag-of-features モデル) 172
bag-of-words model (bag-of-words モデル) 230, 240
bag-of-visual-words model ⇔ bag-of-features model
batch learning (バッチ学習) 13
batch normalization (バッチ正規化) 138
bi-directional recurrent neural network (双方向型回帰結合ニューラルネットワーク) 215
bigram (バイグラム) ⇔ n -gram
binary parse tree (二分構文木) 248
blocked Gibbs sampling (ブロック化ギブスサンプリング) 57
BM ⇔ Boltzmann machine
BoF model ⇔ bag-of-features model
Boltzmann distribution (ボルツマン分布) 39
Boltzmann machine (ボルツマンマシン) 39
 approximation technique of – 54
 deep – 64
 learning of – 42
 learning equation of – 44
 restricted – 60
BoW model ⇔ bag-of-words model
BPTT method ⇔ backpropagation through time method

【C】

CD method ⇔ contrastive divergence method
 CDBN ⇔ convolutional deep belief network
 central difference (中心差分) 150
 cerebral nerve system (脳神経系) 9, 178
 chain rule (連鎖律) 130
 CNN ⇔ convolutional neural network
 combinatorial explosion (組み合わせ爆発) 47
 competitive learning (競合学習) 15
 conditional independence (条件付き独立性) 61
 context (文脈) 233
 continuous bag-of-words model (連続 bag-of-words モデル) 239
 continuous skip-gram model (連続スキップグラムモデル) 241
 contractive autoencoder (縮小自己符号化器) 26
 contrastive divergence (コントラストティブ・ダイバージェンス) 105
 contrastive divergence method (コントラストティブ・ダイバージェンス法) 102
 loss function 103
 persistent – 112
 ConvNet ⇔ convolutional neural network
 convolution (畳み込み) 159
 convolution layer (畳み込み層) 159
 convolutional deep belief network (畳み込み深層信念ネットワーク) 183
 convolutional neural network (畳み込みニューラルネットワーク) 19, 158
 convolution layer 159
 difference from previous methods 172
 extension of – 176
 fully connected layer 159
 internal representation 177
 learning of – 165
 multitask learning 178
 pooling layer 161
 recognition performance 175
 corpus (コーパス) 232
 cost function (コスト関数) ⇔ loss function
 cross entropy (交差エントロピー) 165, 203

【D】

DAE ⇔ deep autoencoder
 dark knowledge (暗黒知識) 141

data parallelism (データ並列化) 132
 DBL method ⇔ detailed balance learning method
 DBM ⇔ deep Boltzmann machine
 DBN ⇔ deep belief network
 decoder (復号化器) 94
 deconvolutional network (逆畳み込みネットワーク) 177
 deep autoencoder (深層自己符号化器) 94
 deep belief network (深層信念ネットワーク) 74
 convolutional – 183
 inference of – 76
 pre-training of – 76
 deep Boltzmann machine (深層ボルツマンマシン) 64
 application of – 73
 learning of – 70
 pre-training of – 66
 deep hierarchical neural network (深層階層型ニューラルネットワーク) ⇔ 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 211
 – hidden Markov model 206
 language model 222
 learning of – 17
 pre-training of – 206
 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) 206
 denoising autoencoder (雑音除去自己符号化器) 25, 117, 208
 dependency parsing (係り受け解析) 229
 detailed balance condition (詳細釣り合い条件) 107
 detailed balance learning method (detailed balance learning 法) 106
 dimension reduction (次元削減) 5, 8
 discriminative pre-training (識別的事前学習) 209
 DistBelief 132
 distillation (蒸留) 140
 distributed parallel computation (分散並列計算) 132

distributed semantic representation (分散の意味表現)	238
distributional hypothesis (分布仮説)	237
distributional memory (分布メモリ)	248
distributional semantic composition (分布の意味構築)	247
distributional semantic representation (分布の意味表現)	238
distributional semantics (分布意味論)	233
divisive normalization (除算正規化)	167
DNN ⇨ deep neural network	
DNN-HMM ⇨ deep neural network-hidden Markov Model	
Downpour SGD	133
DropOut	141
linear regression	144
dropout noise (欠落ノイズ) ⇨ masking noise	
dynamic pooling (動的プーリング)	251

【 E 】

EFH ⇨ exponential family harmonium	
ELM ⇨ extreme learning machine	
Elman network (エルマン・ネットワーク)	215
empirical distribution (経験分布)	45
encoder (符号化器)	94
energy function (エネルギー関数)	38
error correction learning (誤り訂正学習)	12
estimation (推定)	37
exchange Monte Carlo method (交換モンテカルロ法)	113
exponential family harmonium (指数型ハーモニウム族)	98
Gibbs sampling	100
extreme learning machine (極端学習機械)	118

【 F 】

FA ⇨ factor analysis	
factor analysis (因子分析)	8
feature (特徴)	5
feature engineering (特徴工学)	7
feature learning (特徴学習) ⇨ representation learning	
feature point (特徴点)	67

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) ⇨ hierarchical neural network	
Fisher vector (フィッシャーベクトル)	172
fully connected layer (全結合層)	159
function word (機能語)	246

【 G 】

Gabor filter (ガボールフィルタ)	173
Gaussian mixture model (混合正規分布)	197
Gaussian noise (ガウス雑音)	25
Gaussian-Bernoulli restricted Boltzmann machine (ガウス-ベルヌーイ型制限ボルツマンマシン)	206
general object recognition (一般物体認識)	155, 168
previous methods for –	169
generative model (生成モデル)	35
Gibbs distribution (ギブス分布) ⇨ Boltzmann distribution	
Gibbs sampler (ギブスサンプラー) ⇨ Gibbs sampling	
Gibbs sampling (ギブスサンプリング)	55
global feature (大域特徴)	169
GMM ⇨ Gaussian mixture model	
GPU	135
gradient ascent method (勾配上昇法) ⇨ gradient descent method	
gradient descent method (勾配降下法)	47, 129
grandmother cell (おばあさん細胞)	156
graphical model (グラフィカルモデル)	39
greedy layer-wise training (層ごとの貪欲学習)	95, 115
grid search (グリッド探索)	148

【 H 】

harmonium (ハーモニウム) ⇨ restricted Boltzmann machine	
Heaviside function (ヘビサイド関数) ⇨ threshold function	
hidden layer (隠れ層)	11, 60
hidden Markov model (隠れマルコフモデル)	195
deep neural network –	206
neural network –	199
hidden variable (隠れ変数)	42, 52, 96

hierarchical neural network (階層型ニューラルネットワーク)	8, 129, 200
backpropagation method	14
deep – ⇨ deep neural network	
learning of –	11
– model	9
hierarchical softmax (階層型ソフトマックス)	243
hinge loss (ヒンジ損失)	236
HMM ⇨ hidden Markov model	
HNN ⇨ hierarchical neural network	
Hopfield network (ホップフィールド・ネットワーク)	41
hourglass-type neural network (砂時計型ニューラルネットワーク)	22, 93
Huffman tree (ハフマン木)	243
hyper parameter (超パラメータ)	148

【I】

ICA ⇨ independent component analysis	
identity function (恒等関数)	22
ILSVRC	155
image recognition (画像認識)	155
independent component analysis (独立成分分析)	8
inference (推論)	37
InfiniBand	137
input layer (入力層)	11
internal covariate shift (内部共変量シフト)	138
internal layer (中間層) ⇨ hidden layer	
internal representation (内部表現)	5, 17
invariance (不変性)	7, 91, 161, 176, 185, 242
iterative parameter mixing method (反復的パラメータ混合法)	246

【K】

KL distance (KL 距離) ⇨ Kullback-Leibler divergence	
KL divergence (KL ダイバージェンス) ⇨ Kullback-Leibler divergence	
KL information (KL 情報量) ⇨ Kullback-Leibler divergence	
Kullback-Leibler divergence (カルバック-ライブラー・ダイバージェンス)	45, 99

【L】

L-BFGS method (L-BFGS 法)	134, 250
L_1 regularization (L_1 正則化)	25
language model (言語モデル)	195, 232
neural network –	233
large vocabulary continuous speech recognition (大語彙連続音声認識)	198
lasso ⇨ L_1 regularization	
latent representation (潜在表現) ⇨ internal representation	
latent variable (潜在変数) ⇨ hidden variable	
LCA ⇨ local contrast normalization	
learning rate (学習率)	13, 166
tuning of –	147
LeNet ⇨ convolutional neural network	
likelihood function (尤度関数)	43
limited-memory Broyden-Fletcher-Goldfarb-Shanno method ⇨ L-BFGS method	
linear separation (線形分離)	12
linearly separable (線形分離可能)	13
local contrast normalization (局所コントラスト正規化)	166
local feature (局所特徴)	169
local optimum (局所最適解)	15, 87
log mel-filterbank feature vector (対数メルフィ ルタバンク特徴ベクトル)	207
log-bilinear form (対数双線形)	242
log-likelihood function (対数尤度関数)	44
long short-term memory method	218
loss function (損失関数)	86, 129
low effective dimension	148
L_p pooling (L_p プーリング)	162
LSTM method ⇨ long short-term memory method	

【M】

machine learning (機械学習)	4
manifold learning (多様体学習)	8
MAP estimation (MAP 推定) ⇨ maximum a posteriori estimation	
marginalized denoising autoencoder (周辺化雑音除去自己符号化器)	118
Markov chain Monte Carlo method (マルコフ連鎖モンテカルロ法)	55

Markov property (マルコフ性)	55
Markov random field (マルコフ確率場)	38
masking noise (欠落雑音)	25
max pooling (最大プーリング)	162
maximum a posteriori estimation (最大事後確率推定)	37
maximum likelihood estimation (最尤推定)	43
maximum likelihood estimator (最尤推定量)	43
maximum mutual information criterion (最大相互情報量規準)	212
MaxOut	145
mean-field approximation (平均場近似)	58
mean-field equation (平均場方程式)	59
mel-frequency cepstral coefficient (メル周波数ケプストラム係数)	197
MFCC ⇨ mel-frequency cepstral coefficient	
mini batch method (ミニバッチ法)	131
minimum probability flow method (最小確率流法)	110
MLE ⇨ maximum likelihood estimation	
MMI criterion (MMI 規準) ⇨ maximum mutual information criterion	
model bias (モデルバイアス) ⇨ model error	
model error (モデル誤差)	53
model parallelism (モデル並列化)	132
molecular field approximation (分子場近似) ⇨ mean-field approximation	
moment matching (モーメント・マッチング)	45
momentum (モメンタム)	166
MPF method ⇨ minimum probability flow method	
MRF ⇨ Markov random field	
multi-layer neural network (多層ニューラルネットワーク) ⇨ hierarchical neural network	
multi-layer perceptron (多層パーセプトロン) ⇨ hierarchical neural network	
multi-stream learning (マルチストリーム学習)	221
multitask learning (マルチタスク学習) 222, 237	
convolutional neural network	178
mutually connected neural network (相互結合ニューラルネットワーク)	10
MV-RNN ⇨ matrix vector recursive neural network	

【N】

natural gradient method (自然勾配法)	90
natural language processing (自然言語処理)	229
negative sampling (負例サンプリング)	245
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, 158
deep –	16
feed-forward – ⇨ hierarchical neural network	
– –hidden Markov model	199
hierarchical –	8
history of –	9
hourglass-type –	22
multi-layer – ⇨ hierarchical neural network	
mutually connected –	10
recurrent –	215
time-delay –	201
neural network language model (ニューラルネットワーク言語モデル)	233
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)	199
hybrid approach	202
tandem approach	204
neuron (ニューロン)	9
n -gram (n グラム)	198, 230, 232
NLMM ⇨ neural network language model	
NN ⇨ neural network	
NN-HMM ⇨ 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)	39, 96

【O】

objective function (目的関数) ⇨ loss function	
observable variable (観測変数) ⇨ visible variable	
one-versus-rest encoding (一対他符号化) ⇨ 1-of- n representation	
online learning (オンライン学習)	13
output function (出力関数) ⇨ activation function	
output layer (出力層)	11
over-fitting (過学習)	86, 141

【P】

parallel tempering method (パラレル・テンパリング法) ⇔ exchange Monte Carlo method	
paraphrase detection (言い換え表現認識)	248
part-of-speech tagging (品詞タグ付け)	229
partition function (分配関数) ⇔ normalization constant	
PCA ⇔ principal component analysis	
perceptron (パーセプトロン)	9
– learning ⇔ error correction learning	
– learning rule convergence theorem	13
limitation of –	14
multi-layer – ⇔ hierarchical neural network	
perceptron learning (パーセプトロン学習) ⇔ error correction learning	
persistent contrastive divergence method (継続的 コントラストティブ・ダイバージェンス法)	112
phoneme (音素)	194
phonemic context (音素文脈)	196
plateau (プラトー)	87
PoE ⇔ product of experts	
pooling (プーリング)	21, 161, 172
average –	162
dynamic –	251
L_p –	162
max –	162
probabilistic max –	183
pooling layer (プーリング層)	161
post-training (事後学習)	231
posterior distribution (事後分布)	36
pre-training (事前学習)	22, 90, 115, 230
– of deep belief network	76
– of deep Boltzmann machine	66
– of denoising autoencoder	208
discriminative –	209
pre-training of –	115
– by probabilistic models	96
stacked autoencoder	115
principal component analysis (主成分分析)	8
prior distribution (事前分布)	36
probabilistic graphical model (確率的グラフィ カルモデル) ⇔ graphical model	
probabilistic max pooling (確率的最大プー リング)	183
product of experts	120

pronunciation model (発音辞書モデル)	195
pseudo negative instance (擬似負例)	246
pyramid match kernel	172

【R】

random search (ランダム探索)	148
RBM ⇔ restricted Boltzmann machine	
receptive field (受容野)	11, 161
reconstruction error (再構成誤差)	94
reconstruction TICA (再構成型 TICA) ⇔ recon- struction topographic independent component analysis	
reconstruction topographic independent compo- nent analysis (再構成型トポグラフィック独立 成分分析)	184
rectified linear unit ⇔ ReLU	
recurrent neural network (回帰結合ニューラル ネットワーク)	215
acoustic model	215
bi-directional –	215
Elman network	215
language model	223
long short-term memory method	218
recursive autoencoder (再帰自己符号化器)	248
recursive neural network (再帰ニューラルネッ トワーク)	248
recursive autoencoder	250
unfolding recursive autoencoder	147
regret (リグレット)	25
regularization (正則化)	140
distillation	144
DropOut	25
L_1 –	181
sparse –	145, 159
ReLU	7, 180, 230
representation learning (表現学習)	60, 96, 206
restricted Boltzmann machine (制限ボルツマン マシン)	72
– as autoencoder	60
Bernoulli-Bernoulli –	61
conditional independence of –	206
Gaussian-Bernoulli –	62
learning of –	
RNN ⇔ recurrent neural network, ⇔ recursive neural network	

robustness (ロバスト性)	7
【 S 】	
saddle point (鞍点)	87
salt-and-pepper noise (ごま塩雑音)	25
Sandblaster L-BFGS	133
scale invariant feature transform	☞ SIFT
semantic representation (意味表現)	236, 237
sequence discriminative training (系列識別学習)	211
SGD	☞ stochastic gradient descent method
shared weight (共有重み)	161
SIFT	170
sigmoid belief (シグモイド信念)	41
sigmoid function (シグモイド関数)	14
simple perceptron (単純パーセプトロン)	12
smoothing (平滑化)	233
softmax function (ソフトマックス関数)	159
– output layer	203
sparse autoencoder (スパース自己符号化器)	24, 181
sparseness (スパース性)	8
speaker adaptation (話者適応)	204
speech recognition (音声認識)	195
stacked autoencoder (積層自己符号化器)	22, 72 , 115
statistical machine learning (統計的機械学習)	35
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 , 131
stop word (不用語)	246
subtractive normalization (減算正規化)	166
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)	113
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)	201

topographic independent component analysis (トポグラフィック独立成分分析)	184
reconstruction –	184
topographic mapping (トポグラフィック・マッピング)	185
transfer learning (転移学習)	☞ multitask learning
trigram (トライグラム)	☞ <i>n</i> -gram

【 U 】

ugly duckling theorem (醜いアヒルの子の定理)	6
undirected graph (無向グラフ)	38
unfolding recursive autoencoder (展開再帰自己符号化器)	250
unigram (ユニグラム)	☞ <i>n</i> -gram
unsupervised learning (教師なし学習)	12

【 V 】

vanishing gradient problem (勾配消失問題)	18 , 218
variational autoencoder (変分自己符号化器)	114
variational Bayes method (変分ベイズ法)	☞ mean-field approximation
visible layer (可視層)	60
visible variable (可視変数)	42, 96
visual word (ビジュアルワード)	170

【 W 】

weight (重み)	10
weight decay (重み減衰)	166
weight sharing (重み共有)	161
whitening (白色化)	167
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)	211
word2vec	239

【 Z 】

zero-frequency problem (ゼロ頻度問題)	198, 233
---------------------------------	-----------------