

構造適応型 Deep Belief Network からの 知識獲得に関する一考察

A Consideration of Knowledge Acquisition from Adaptive Learning Method of Deep Belief Network

鎌田 真

Shin Kamada

広島市立大学大学院情報科学研究科情報科学専攻

Email: da65002@e.hiroshima-cu.ac.jp

市村 匠

Takumi Ichimura

県立広島大学経営情報学部経営情報学科

Email: ichimura @ pu-hiroshima.ac.jp

Abstract—Deep Belief Network (DBN) has an deep architecture that can represent multiple features of input patterns hierarchically with pre-trained Restricted Boltzmann Machines (RBM). The model of DBN has an advantage of visualization or knowledge acquisition of the trained network because it is generative stochastic model. We have already proposed the adaptive learning method of DBN that can find an optimal number of hidden neurons and layers in the learning phase. In this paper, some considerations about the knowledge acquisition of the trained network by visualizing the activation of hidden neurons is discussed.

I. はじめに

近時、人工知能の研究が産業において具体的にかつ活発に応用されるようになってきた。特に深層学習として知られる Deep Learning は、ニューラルネットワークを用いた自然言語処理、音声処理、画像処理などにおいて、高い分類能力をもつ学習方法として注目を集めている [1], [2]。

Deep Belief Nets(DBN)[3] は、生成モデルに基づく Deep Learning の学習手法である。統計的な手法である Restricted Boltzmann Machine(RBM)[4] を用いて入力データが持つ分布(事前知識)を各層ごとに学習し、これらを組み合わせることで入力データに含まれる一般的な特徴から具体的で詳細な特徴を表現することができる。他の Deep Learning の学習手法としては、Auto Encoder を階層的に組み合わせた Stacked Auto Encoder[5] や畳み込みニューラルネットワーク [6] が知られている。これらの手法が決定論的なモデルであるのに対して、DBN は統計的な手法に基づいた生成モデルであるため、学習したモデルの最適性を誤差だけでなく尤度の観点からも議論できる。また、学習後のネットワークの可視化に優れていると考えられる。

Deep Learning は多くの分野で優れた成果を残しつつも、(1) 最適なネットワーク構造は入力データの分布に応じて異なり、パラメタ設定は試行錯誤的に求め

る必要がある、(2) 学習後のネットワーク構造はブラックボックスであり、学習結果の利活用、すなわち知識獲得が困難であることが知られている。我々の研究では、これらの問題を解決するために、DBN において最適な隠れニューロン数および層の数を学習中に自動で求める構造適応型学習手法を提案した [7]。さらに、ネットワークの重みの正則化や隠れニューロンの出力をスパースにすることで、出力値の意味付けを行う忘却学習 [8] を DBN に適用し、有効性を示した。これらの手法をいくつかの画像ベンチマークデータセットに適用したところ、従来の RBM や DBN よりも高い分類精度が得られた [9], [10]。本論文では、DBN の構造適応型学習により得られたネットワーク構造から知識獲得を行うことを試みた。このために、入力データを与えたときに、出力に至るまでに発火する隠れニューロンのパスを可視化する描画するシステムを開発し、知識獲得を試みる。

II. 構造適応型 RESTRICTED BOLTZMANN MACHINE による学習手法

A. Restricted Boltzmann Machine

RBM[4] は、可視層と隠れ層の 2 層から構成されるネットワーク構造を持ち、確率分布に基づいて入力データに含まれる特徴を隠れニューロン上で学習することができる。従来の Boltzmann Machine[11] では、各層のニューロン同士に相互結合があるが、RBM では、図 1 のように各層のニューロン間の結合はない。このため、計算量を削減するだけでなく、隠れ層の各ニューロンごとに独立した確率分布を学習することができる。

RBM の学習では、エネルギー関数を最小にするためのパラメタが最尤推定により求められる。ここで、 v_i を i 番目の可視ニューロン、 h_j を j 番目の隠れニューロンとしたとき、エネルギー関数 $E(\mathbf{v}, \mathbf{h})$ は式 (1) のように定義され、また入力 \mathbf{v} と \mathbf{h} の尤度 $p(\mathbf{v}, \mathbf{h})$ は式 (2) のように定義される。ここで、 $\mathbf{v} \in \{0, 1\}^I$ と $\mathbf{h} \in \{0, 1\}^J$

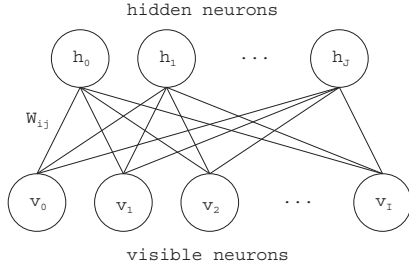


図 1. Structure of RBM

はそれぞれ可視層と隠れ層のベクトルであり、 I と J は可視ニューロンの数、隠れニューロンの数である。

$$E(\mathbf{v}, \mathbf{h}) = \sum_i b_i v_i - \sum_j c_j h_j - \sum_i \sum_j v_i W_{ij} h_j, \quad (1)$$

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} \exp(-E(\mathbf{v}, \mathbf{h})), \quad (2)$$

$$Z = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h})), \quad (3)$$

ここで、 b_i は v_i に対するパラメタ、 c_j は h_j に対するパラメタ、 W_{ij} は v_i と h_j 間の重みである。また Z は分配関数 (partition function) であり、 \mathbf{v} と \mathbf{h} の全ての可能な組み合わせに対するエネルギーを示す。RBM の学習では、入力データの確率分布を表現するためのパラメタ $\theta = \{\mathbf{b}, \mathbf{c}, \mathbf{W}\}$ を最尤推定により求めることができる。ただし、計算量の観点から厳密な解を求めることは困難であるため、一般的には、サンプリングに基づいた Contrastive Divergence(CD) 法 [12] による近似が行われる。CD 法における最適なサンプリング回数については議論がされているが、サンプリング数が 1 回 (CD-1) でも性能が高いことが知られている [13]。

B. ニューロン生成／消滅アルゴリズムによる構造適応型学習

我々は、階層型ニューラルネットワークに対して開発された文献 [14] の手法をもとに、RBM の学習中に最適な隠れニューロン数を自動で求めるニューロン生成／消滅アルゴリズムによる構造適応型学習手法を提案している [9]。階層型ニューラルネットワークでは、学習が進むにつれて重みベクトルの変分が収束すること (図 2) を基本とし、ニューロンに過不足があれば、この過程で局所解に陥ることをもとにニューロンの生成／消滅を行っている。さらに、自動生成したネットワーク構造の最適性を凸性および連続性を用いて証明しているが、RBM の学習では、CD 法によるサンプリングの過程で、条件付き確率である連続値が 2 値に変換されてしまうため、連続性を満たさない場合があると我々は考えている。これに対して、先行研究では、学習における各パラメタの変分を一定の範囲内に収めることで、学習が収束することをリップシッツ連続を用いて証明している [15]。一方で、この条件を満たさず、パラメタの変分が収束しない場合は、学習データに対

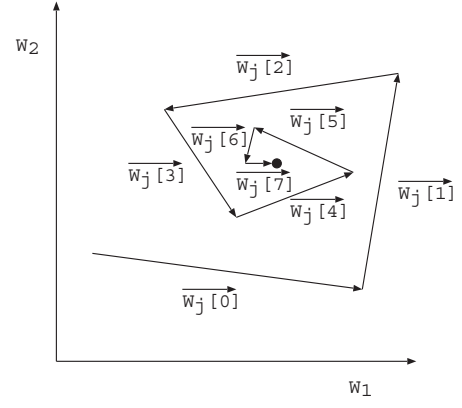


図 2. Convergence situation of a weight vector

するネットワークの表現能力が不足しており、データが持つ特徴 (確率分布) を隠れニューロン上で線形分離できていないことが原因だと考えられる。このような場合、関連する位置に十分な数の隠れニューロン挿入することで、連続性を満たさないまでも、多くの次元でデータを表現できると考えられ、ネットワークが安定し、収束する。我々の調査では、RBM の 3 つのパラメタ $\theta = \{\mathbf{b}, \mathbf{c}, \mathbf{W}\}$ のうち、 \mathbf{c} と \mathbf{W} が CD 法による学習の収束に関係していることを発見し [16]、ニューロン生成の条件を式 (4) に示すように定義した。

$$(\alpha_c \cdot dc_j) \cdot (\alpha_W \cdot dW_{ij}) > \theta_G, \quad (4)$$

ここで、 dc_j 、 dW_{ij} は、それぞれ j 番目の隠れニューロンのパラメタの勾配、 i, j 番目の重みの勾配を示し、 α_c 、 α_W は、各パラメタの勾配のスケールを調整するために使われるパラメタであり、 θ_G は閾値である。学習中に式 (4) を満たす隠れニューロンがある場合、図 3(a) のように、該当する隠れニューロンの近傍に新しいニューロンがその属性 (パラメタや重みの値) を継承するように挿入される。

一方、入力データに対して十分な数の隠れニューロンが生成された後、その中に出力に寄与していない、すなわち冗長な値を出力する隠れニューロンが存在する場合がある。このようなニューロンがある場合、学習に係る計算コストが増加するだけでなく、ネットワークの構造が複雑になり、誤った出力を生じる場合がある。提案する構造適応型 RBM では、一定の学習の後、式 (5) を満たす隠れニューロンが存在すれば、ニューロン消滅アルゴリズムを適用し、該当する隠れニューロンを図 3(b) のように消去する。式 (4) では、すべての入力データに対する隠れニューロンの出力値の平均をモニタリングし、その値が一定の閾値以下のニューロンを冗長なニューロンとみなし、削除している。

$$\frac{1}{N} \sum_{n=1}^N p(h_j = 1 | \mathbf{v}_n) < \theta_A, \quad (5)$$

$$p(h_j = 1 | \mathbf{v}_n) = \text{sigm}(b_j + \sum_i W_{ij} v_i), \quad (6)$$

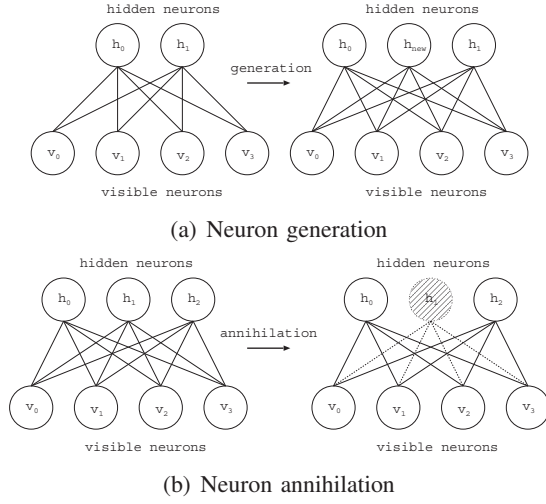


図 3. Adaptive Learning method of RBM

ここで、 $v_n = \{v_1, v_2, \dots, v_N\}$ は与えられた入力データセットであり、 N はそのレコード数である。 $p(h_j = 1|v_n)$ は RBM の可視層に入力データ v_n が与えられたときの隠れニューロン $h_j \in \{0, 1\}$ の発火確率を示している。ここで、 $\text{sigm}()$ はシグモイド関数で $[0, 1]$ を出力する関数である。 θ_A はしきい値である。

C. 忘却機能を持った構造学習

RBM によるニューロン生成／消滅アルゴリズムにより、入力データ空間に応じて最適なネットワーク構造を学習中に求めることができるが、学習後のネットワーク構造はブラックボックスである。従って、ネットワークに蓄積された重みや隠れニューロンの出力値と、入力データとの関係性が不明確であり、そこから IF-THEN ルールで表現されるような明示的な知識の獲得は困難である。このような問題を解決するため、石川は、階層型ニューラルネットワークにおいて、ネットワークから知識獲得を行うための忘却付き構造学習法を提案している [8]。ここでは、“忘却学習”、“隠れユニット明確化”、“選択的忘却学習”の 3 つの手続きによりネットワークの正則化を行い、スパースな構造をもつニューラルネットワークを構成している。我々は、この手法に基づいた RBM の忘却学習手法を開発し、文献 [10], [17] においてその有効性を示している。

式 (7) から式 (9) は、忘却学習による 3 種類の目的関数を示しており、それぞれ忘却学習、隠れユニット明確化、選択的忘却学習を示している。

$$J_f = J + \sum_{i,j} |w_{ij}|, \quad (7)$$

$$J_h = J + \sum_i \min\{1 - h_i, h_i\}, \quad (8)$$

$$J_s = J + \sum_{|w_{ij}| < \theta} |w_{ij}|, \quad (9)$$

ここで、 J は学習における通常の目的関数であり、RBM の場合は CD 法の推定により生じる誤差を示している。

式 (7) の忘却学習では、重みの絶対値の和を目的関数に加えている。これは、L1 正則化の効果があり、重みが必要以上に大きくなることを抑制している。式 (8) の隠れユニット明確化では、隠れユニットにおける出力を強制的に 2 値化するように評価する項を追加している。式 (9) の選択的忘却学習では、式 (7) の忘却学習によって目的関数の値が通常より大きくなってしまふことを避けるために、しきい値の範囲内の重みに限定して忘却学習を行っている。これらの式を各パラメタに対して偏微分することで学習におけるパラメタの修正量を求めることができる。RBM では、式 (1) に示すようにパラメタ c と W が隠れニューロン h に関連していることから、式 (10) および式 (11) の右辺の第 3 項が忘却学習による勾配として、通常の CD 法におけるパラメタの更新式に追加される。

$$\Delta w_{ij} = v_i p(h_j = 1|v) - v'_i p(h'_j = 1|v') - \frac{\lambda_1}{2} \text{sgn}(w_{ij}), \quad (10)$$

$$\Delta c_j = p(h_j = 1|v) - p(h'_j = 1|v') - \frac{\lambda_2}{2} (1 - p(h_j = 1|v)), \quad (11)$$

ここで、 λ_1 と λ_2 はそれぞれ重みと隠れニューロンの修正量を調整するパラメタである。文献 [8] で提唱されているように、忘却学習と隠れユニットの明確化は学習中に同時に適用すべきである (選択的忘却学習は、最終的な誤差を小さくするために学習の後半に適用する)。我々の調査では、RBM の学習において λ_1 と λ_2 を異なる値に設定した場合、 W と c の収束状況において、片方が学習早期に収束する現象が見られたため、同じ値 ($\lambda_1 = \lambda_2 = 1$) とした。

III. 構造適応型 DEEP BELIEF NETWORK による学習手法

Deep Belief Network (DBN) は、Hinton が提案した Deep Learning の手法であり、事前学習した RBM を複数積み重ねることで、入力データの特徴を階層的に学習することができる [3]。図 4 は、3 つの RBM を持つ DBN のネットワーク構造を示している。入力層を 0 層とした際、 $l (> 0)$ 層のエネルギーは式 (12) のように定義される。また、 l 層の j 番目の隠れニューロンの発火確率は式 (13) のように定義される。

$$E^l = E(\mathbf{h}^{l-1}, \mathbf{h}^l) = (\mathbf{b}^l)^T \mathbf{h}^{l-1} - (\mathbf{c}^l)^T \mathbf{h}^l - \mathbf{h}^{l-1} \mathbf{W}^l \mathbf{h}^l, \quad (12)$$

$$p(h_j^l = 1|\mathbf{h}^{l-1}) = \text{sigm}(\mathbf{c}^l + \mathbf{W}^l \mathbf{h}^{l-1}), \quad (13)$$

ここで、 \mathbf{b}^l 、 \mathbf{c}^l はそれぞれ l 層の RBM の可視ニューロンと隠れニューロンのパラメタであり、 \mathbf{W}^l はその重みである。 $\mathbf{h}^l \in \{0, 1\}$ は隠れニューロンの状態ベクトルである。

本論文では、DBN の学習中において、入力データに対して最適な隠れ層の数を自動で求める構造適応型学

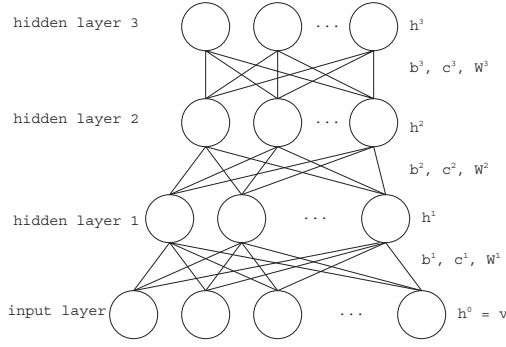


図 4. Hierarchical network structure of DBN

表 I
CLASSIFICATION ACCURACY

	Training	Test
Traditional RBM [20]	-	63.0%
Traditional DBN [21]	-	78.9%
Adaptive RBM	99.9%	81.2%
Adaptive RBM with Forgetting	99.9%	85.8%
Adaptive DBN with Forgetting	100.0%	92.4%

習手法を提案する。構造適応型 RBM におけるニューロン生成アルゴリズムでは、学習中の WD (パラメタ c と W の変分) をモニタリングしたが、提案する構造適応型 DBN では、個々の RBM の安定性ではなく、ネットワーク全体の安定性を測るため、式 (14) を用いてネットワーク全体の WD およびエネルギーを観察した。

$$\sum_{l=1}^k (\alpha_{WD} \cdot WD^l + \alpha_E \cdot E^l) > \theta_L, \quad (14)$$

ここで、 WD^l は l 層の RBM における c^l と W^l の変分を示している。 E^l はエネルギー関数である。 α_{WD} と α_E はスケールを調整するために使われるパラメタであり、 θ_L は閾値である。ある k 層の RBM の学習中に式 (14) を満たす場合、入力データに対するネットワーク全体の表現能力が不足していると考え、 k 層の学習が終了した時点で、 $k+1$ 層の RBM を生成し、学習を行う。ここでも RBM のニューロン生成アルゴリズムと同様に、 $k+1$ 層の RBM のパラメタ b , c , W の初期値が k 層の RBM から継承される。

表 II
SITUATION IN EACH LAYER AFTER TRAINING OF ADAPTIVE DBN

Layer	No. neurons	Total energy	Total error	Accuracy
1	433	-0.24	25.37	84.6%
2	1595	-1.01	10.76	86.2%
3	369	-0.78	1.77	90.6%
4	1462	-1.00	0.43	92.3%
5	192	-1.17	0.01	92.4%

IV. 実験

A. データセット

提案する構造適応型 DBN の性能を評価するために、CIFAR-10[18] と呼ばれるカラー画像データセットを用いた識別実験を行った。CIFAR-10 は、50,000 枚の訓練画像と 10,000 枚のテスト画像から構成され、訓練画像とテスト画像間に重複する画像はない。これらの画像は車、犬などの 10 種類のクラスを持ち、各画像は 32×32 ピクセルである。

RBM, DBN の実装には、Pylearn2[19] と呼ばれる Deep Learning のライブラリを用いた。学習アルゴリズムには Stochastic Gradient Descent (SGD) を使い、学習率は 0.1、学習のバッチサイズは 100 とした。構造適応型 RBM のパラメタとして、 $\theta_G = 0.015$, $\theta_A = 0.0005$ とした。構造適応型 DBN のパラメタとして、 $\theta_L = 0.1$ とした。

B. 実験結果

表 I は、CIFAR-10 の訓練事例とテスト事例に対する識別精度を示している。従来の RBM[20] や DBN[21] に比べて、提案する構造適応型 RBM は高い識別精度を示し、さらに、忘却学習を適用することでテスト事例に対する識別精度が向上した。この中でも、本論文で提案する構造適応型 DBN は訓練事例だけでなくテスト事例に対して最も高い分類精度を示すことができた。また、表 II は、構造適応型 DBN における各層の学習状況を示している。上位層に進むにつれて、エネルギーや誤差は小さくなり、識別精度は向上した。4 層から 5 層にかけて、エネルギーや誤差の変化が少なくなり、式 (14) による層の自動生成は 5 層で終了した。

ここで、誤答した事例の中には、猫を犬と判別するなど、画像の中に複雑であいまいな特徴が含まれていることが分かった。そこで、このようにあいまいな特徴を持つ事例の判別や学習したネットワークの知識獲得を行うための調査として、テスト事例における各クラスに対して正答した事例と誤答した事例をそれぞれ学習済みのネットワークに入力として与え、出力に至るまで各層において発火した隠れニューロンのパスを可視化した。図 5(a) から図 5(d) は、CIFAR-10 のクラスの猫と犬に対して、正答した事例と誤答した事例を与えた場合のネットワークのパスを示している。ここで、ノードの中の数字は隠れニューロンの添字を示している。ノード間のアークは重みを示し、重みの強さに応じてアークを太字で表している (5 段階)。また、誤答した事例である図 5(b) と図 5(d) において、正答した事例では発火しなかったノードを太字で示している。

同じクラスにおいて正答した事例と誤答した事例で発火したニューロンのパスを比較すると、全体的に、下位層では比較的同じ位置の隠れニューロンが発火したが、上位層に進むにつれてこの数は減少し、誤答した事例は、正答した事例と異なるパスを通っていることが分かった。例えば、図 5(a), 図 5(b) に示すように、1 層では 221 番, 281 番, 2 層では 15 番, 1594 番の

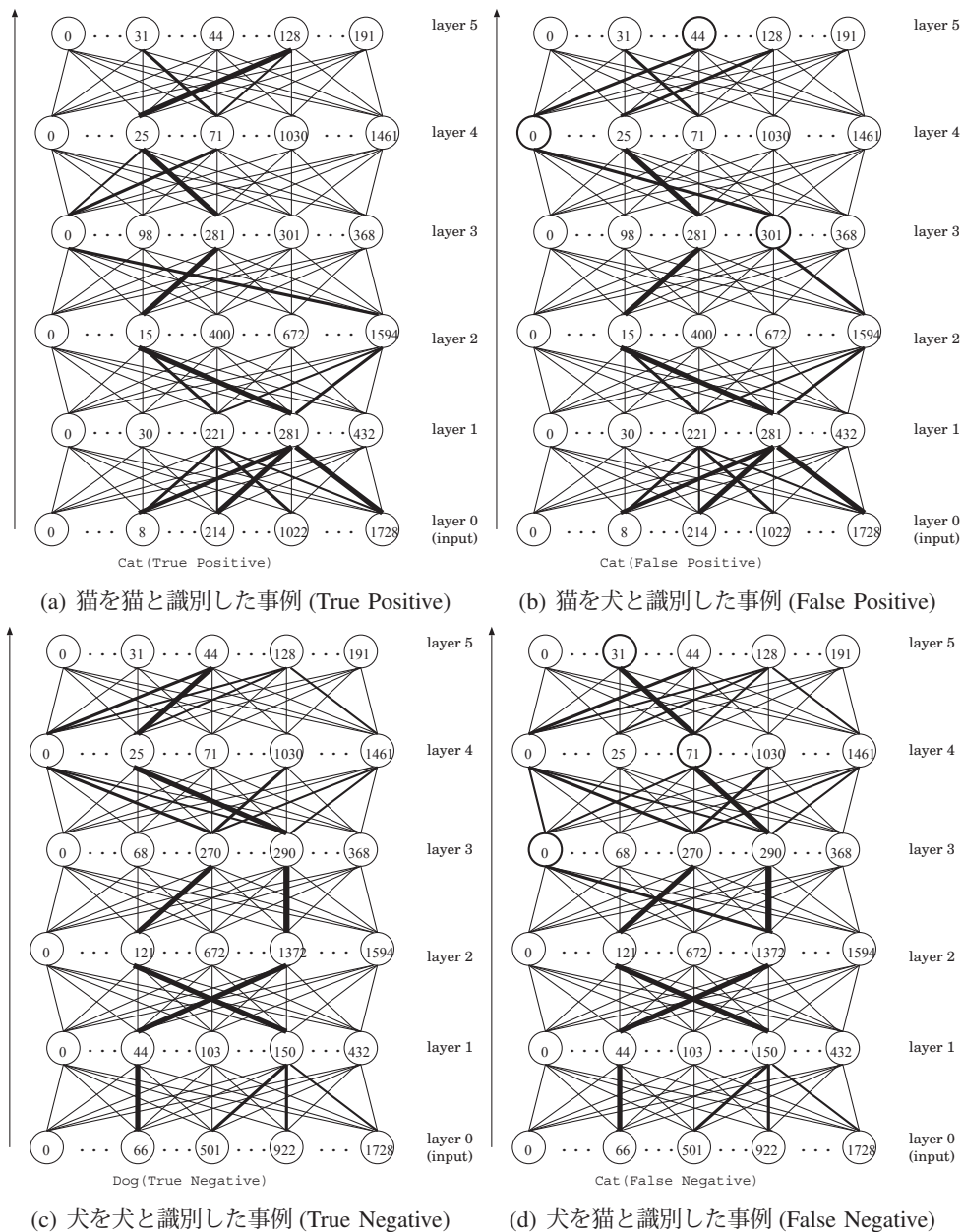


図 5. 学習後の各層の隠れニューロンの発火パス

ニューロンを通るパスの重みがどちらの事例においても強くなった。3層において、正答した事例では0番、281番、301番を通るパスの重みが強くなったが、誤答した事例では0番を通る重みが弱くなり、代わりに、301番を通る重みが強くなった。これに伴い、以降の層において発火する隠れニューロンに変化が現れた。また、図 5(b) と図 5(c) および図 5(a) と図 5(d) において、同じニューロンを発火するケースが見られた。猫を犬と識別した事例である図 5(b) では、5層において44番のノードが誤って発火しているが、犬を犬と識別した事例である図 5(c) では、このノードが発火している。同じように、猫を猫と識別した事例である図 5(a) では、5層において31番のノードが発火しているが、犬を猫と識別した事例である図 5(c) でも、このノードが発火している。これらは、犬や猫を区別するためのニューロンであると考えられる。さらに、図 5(a) から図 5(d)

の全てにおいて、5層の128番の隠れニューロンが発火し、これは猫と犬の共通した特徴を持つニューロンであると考えられる。

このように、誤って識別した事例では、正答した事例に含まれる特徴を表すニューロンだけでなく、別のクラスの特徴を表すニューロンも上位層で同時に発火していた。例えば、猫を犬と識別した例では、5層において、猫の特徴を表す31番のニューロンだけでなく、犬の特徴を表す44番のニューロンも同時に発火している(図 5(b))。その原因は、3層以降において、正答した事例とは異なるパスを通ってしまったことである。この結果を活用することで、発火するパスのパターンからデータの特徴に応じたクラスターリングが可能になる。例えば、犬と猫で誤って識別した事例のように、複数の特徴を持ったあいまいな事例のみを分類することができる。

V. おわりに

Deep Learning と呼ばれる機械学習の手法が様々な分野で高い精度を示しており、現実世界の問題に応用されている。これらは特に画像認識の分野で優れた成果を残しているが、入力データ空間に応じたネットワーク構造の最適化や学習結果の可視化の課題は解決されていない。本研究では、尤度の概念を用いた統計的なモデルであるため、構造の最適性の証明や学習結果の可視化に優れている RBM, DBN に着目し、学習中にニューロンの生成／消滅を行うことで最適な構造を求める構造適応型 DBN を開発した。本論文では、学習後のネットワークの知識獲得に着目し、各層において発火する隠れニューロンのパスを可視化した。調査したところ、誤答した事例は、上位層において異なる隠れニューロンが発火していることが分かり、そのパスにおいて分岐している箇所を発見した。今後は、得られた発火パスのパターンから、IF-THEN ルールのような明示的な知識を獲得する手法を開発する。

参考文献

- [1] Y.Bengio, *Learning Deep Architectures for AI*. Foundations and Trends in Machine Learning archive, Vol.2, No.1, pp.1–127 (2009)
- [2] V.Le.Quoc, R.Marc’s Aurelio, et.al, *Building high-level features using large scale unsupervised learning*, International Conference in Machine Learning (2012)
- [3] G.E.Hinton, S.Osindero and Y.Teh, *A fast learning algorithm for deep belief nets*. Neural Computation, Vol.18, No.7, pp.1527–1554 (2006)
- [4] G.E.Hinton, *A Practical Guide to Training Restricted Boltzmann Machines*. Neural Networks, Tricks of the Trade, Lecture Notes in Computer Science, Vol.7700, pp.599–619 (2012)
- [5] Y.Bengio, P.Lamblin, D.Popovici and H.Larochelle, *Greedy Layer-Wise Training of Deep Networks*, in Advances in Neural Information Processing Systems 19 (NIPS ‘06), pp.153–160 (2007)
- [6] Y.LeCun, et.al., *Backpropagation applied to handwritten zip code recognition*. Neural Computation, Vol.1, No.4, pp.541–551 (1989)
- [7] 鎌田真, 市村匠, 原章, “ニューロン生成／消滅アルゴリズムによる構造適応型 Restricted Boltzmann Machine”, 計測自動制御学会第 8 回コンピューテーション・インテリジェンス研究会, pp.90–96 (2015)
- [8] M.Ishikawa, *Structural Learning with Forgetting*. Neural Networks, Vol.9, No.3, pp.509–521 (1996)
- [9] S.Kamada and T.Ichimura, *An Adaptive Learning Method of Restricted Boltzmann Machine by Neuron Generation and Annihilation Algorithm*. Proc. of IEEE SMC 2016, (to appear in 2016)
- [10] S.Kamada and T.Ichimura, *A Structural Learning Method of Restricted Boltzmann Machine by Neuron Generation and Annihilation Algorithm*. ICONIP 2016, (submitted)
- [11] D.H.Ackley, G.E.Hinton and T.J.Sejnowski, *A Learning Algorithm for Boltzmann Machines*, Cognitive Science, 9, 147–169. doi, 10.1207/s15516709cog0901_7 (1985)
- [12] G.E.Hinton, *Training products of experts by minimizing contrastive divergence*. Neural Computation, Vol.14, pp.1771–1800 (2002)
- [13] T.Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient*, Proc. of the 25th international conference on Machine learning, pp.1064–1071 (2008)
- [14] T.Ichimura and K.Yoshida Eds., *Knowledge-Based Intelligent Systems for Health Care*. Advanced Knowledge International (ISBN 0-9751004-4-0) (2004)
- [15] D.Carlson, V.Cevher and L.Carlin, *Stochastic Spectral Descent for Restricted Boltzmann Machines*. Proc. of the Eighteenth International Conference on Artificial Intelligence and Statistics, pp.111–119 (2015)
- [16] S.Kamada and T.Ichimura, *A Learning Method of Adaptive Deep Belief Network by using Neuron Generation and Annihilation Algorithm*. Proc. of 17th Annual Meeting of Self-Organizing Maps in Japanese, pp.12.1–6 (2016)
- [17] S.Kamada, Y.Fujii and T.Ichimura, *Structural Learning Method of Restricted Boltzmann Machine with Forgetting*. Proc. of 17th Annual Meeting of Self-Organizing Maps in Japanese, pp.13.1–6 (2016)
- [18] A.Krizhevsky, *Learning Multiple Layers of Features from Tiny Images*. Master of thesis, University of Toronto (2009)
- [19] I.Goodfellow, David Warde-Farley, et.al., *Pylearn2: a machine learning research library*, arXiv preprint arXiv:1308.4214 (2013)
- [20] S.Dieleman and B.Schrauwen, *Accelerating sparse restricted Boltzmann machine training using non-Gaussianity measures*. Deep Learning and Unsupervised Feature Learning (NIPS-2012) (2012)
- [21] A.Krizhevsky, *A Convolutional, A Convolutional Deep Belief Networks on CIFAR-10*, Technical report (2010)

問い合わせ先

〒 734-8558

広島県広島市南区宇品東一丁目 1 番 7 1 号

県立広島大学経営情報学部

市村 匠