

確率的極端学習機械による制限ボルツマンマシン分類器の多層化

菅野 友理[†]山形大学大学院理工学研究科[†]安田宗樹[‡]山形大学大学院理工学研究科[‡]

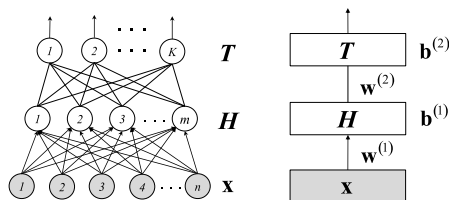
1. はじめに

人工知能を実現する技術の一つに機械学習があり、その中でも一般に4層以上の構造を持つモデルを用いるものを深層学習[1]と呼ぶ。また、深層学習の問題のひとつに、入力データをあらかじめ設定したクラス(カテゴリ)に振り分ける問題である多クラス分類問題(パターン認識)がある。この際、分類するために用いられる確率モデルは分類器(classifier)と呼ばれる。分類器モデルの多層化拡張は、学習性能の向上を目的とした深層学習の枠組みの上で非常に重要な要素であるが、モデル構造が複雑になることから学習達成が困難になる場合が多い。

そこで本稿では、制限ボルツマンマシン分類器(discriminative restricted Boltzmann machine; DRBM)[2]と呼ばれる分類器モデルの、提案モデルである確率的極端学習機械(Probabilistic Extreme Learning Machine; PELM)を用いた多層化を試みる。PELMは、極端学習機械(Extreme Learning Machine; ELM)[3]における非学習型パラメータの考え方を基に、素子を取る値の確率化を導入したモデルである。この提案法を用いて実データを用いた認識実験を行い、多層化による認識性能への影響を確認する。

2. 制限ボルツマンマシン分類器

DRBMは制限ボルツマンマシン(restricted Boltzmann machine; RBM)と呼ばれるマルコフ確率場(Markov random field)型の確率モデルが基になっている3層構造の分類器モデルである。

図1 DRBM. 入力層 $\mathbf{x}$, 中間層 $\mathbf{H}$, 出力層 $\mathbf{T}$ の3層からなる。

DRBMは図1のような入力層 $\mathbf{x}$, 中間層 $\mathbf{H}$, 出力層 $\mathbf{T}$ の3層からなる構造を持つ。中間層と出力層の間のリンクは無向リンク、中間層と入力層の間のリンクは下層から上層への有向リンクになっている。 $\mathbf{b}^{(1)}$, $\mathbf{b}^{(2)}$ は中間層、出力層の素子にかかるバイアスパラメータであり、 $\mathbf{w}^{(1)}$, $\mathbf{w}^{(2)}$ はそれぞれ入力層-中間層、中間層-出力層のリンクに割り当てられている結合重みパラメータである。また、入力層、中間層、出力層はそれぞれ、入力ベクトル $\mathbf{x} = \{x_i \in (-\infty, +\infty) \mid i = 1, 2, \dots, n\}$, 中間

ベクトル $\mathbf{h} = \{h_j \in \{-1, +1\} \mid j = 1, 2, \dots, m\}$ 出力ベクトル $\mathbf{t} = \{t_k \in \{0, 1\} \mid k = 1, 2, \dots, K; \sum_{k=1}^K t_k = 1\}$ を実現値として取る。

これらを用いて DRBM は次の条件付き確率としてモデル化される。

$$P(\mathbf{t}, \mathbf{h} \mid \mathbf{x}, \theta) = \frac{1}{Z(\mathbf{x}, \theta)} \exp \left(\sum_{k=1}^K b_k^{(2)} t_k + \sum_{j=1}^m b_j^{(1)} h_j + \sum_{k=1}^K \sum_{j=1}^m w_{k,j}^{(2)} t_k h_j + \sum_{j=1}^m \sum_{i=1}^n w_{j,i}^{(1)} h_j x_i \right) \quad (1)$$

ここで、 $Z(\mathbf{x}, \theta)$ は規格化条件を満たすための規格化定数である。また、パラメータの集合として $\theta = \{\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \mathbf{w}^{(1)}, \mathbf{w}^{(2)}\}$ としている。

さらに、DRBMにおいて、中間層 $\mathbf{H}$ は出力の計算に余計であるため、周辺化(marginalization)を用いて次のように確率変数を消去する。

$$P(\mathbf{t} \mid \mathbf{x}, \theta) = \sum_{\mathbf{h}} P(\mathbf{t}, \mathbf{h} \mid \mathbf{x}, \theta) \quad (2)$$

ここで、正解と入力のセットが N 個ある訓練データ集合 $D = \{(\mathbf{x}^{(\mu)}, \mathbf{t}^{(\mu)}) \mid \mu = 1, 2, \dots, N\}$ を用いた DRBM の学習は、対数尤度関数

$$l_D(\theta) = \frac{1}{N} \sum_{\mu=1}^N \ln P(\mathbf{t}^{(\mu)} \mid \mathbf{x}^{(\mu)}, \theta) \quad (3)$$

の θ に関する最大化によって達成される。

3. 確率的極端学習機械を用いた制限ボルツマンマシン分類器の多層化

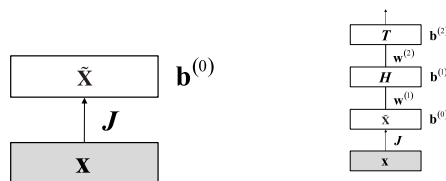


図2 PELM.

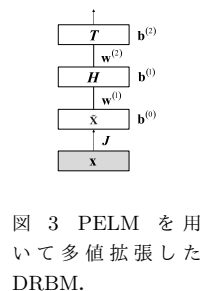


図3 PELM を用いて多値拡張した DRBM.

PELM は、ELM における非学習型パラメータによって入力を変換するという考え方を基に、さらにその素子を取る値が確率的に決定されるモデルである。PELM は、図2のように、入力層の上層に追加される。この新たな中間層 $\tilde{\mathbf{x}}$ は $\tilde{\mathbf{x}} = \{\tilde{x}_r \in \{-1, +1\} \mid r = 1, 2, \dots, \tilde{m}\}$ の実現値を取る。

さらに、入力層 $\mathbf{x}$ と中間層 $\tilde{\mathbf{x}}$ 間の重みパラメータを $\mathbf{J}$, $\tilde{\mathbf{x}}$ にかかるバイアスパラメータを $\mathbf{b}^{(0)}$ とする。このとき、 $\tilde{\mathbf{x}}$ の条件付き

Multilayered discriminative restricted Boltzmann machine with probabilistic extreme learning machine

[†] Yuri Kanno; Graduate School of Science and Engineering, Yamagata University

[‡] Muneki Yasuda; Graduate School of Science and Engineering, Yamagata University

確率を次のように定義する.

$$P(\tilde{\mathbf{x}} | \mathbf{x}, \mathbf{b}^{(0)}, \mathbf{J}) = \prod_{r=1}^{\tilde{m}} P_r(\tilde{x}_r | \mathbf{x}, \mathbf{b}^{(0)}, \mathbf{J}) \quad (4)$$

$$P_r(\tilde{x}_r | \mathbf{x}, \mathbf{b}^{(0)}, \mathbf{J}) = \frac{1}{\tilde{Z}_r} \exp \left(b_r^{(0)} + \sum_{i=1}^n J_{r,i} x_i \tilde{x}_r \right),$$

$$\tilde{Z}_r = 2 \cosh \left(b_r^{(0)} + \sum_{i=1}^n J_{r,i} x_i \right) \quad (5)$$

これにより計算された確率 P_r にしたがって, $\tilde{x}_r$ は -1 か 1 がどちらかの値を確率的に取る.

PELM を用いて DRBM を拡張すると, 図 3 のような形になる. 条件付き確率は $\tilde{\mathbf{x}}$ の期待値を用いて

$$P(\mathbf{t} | \mathbf{x}, \theta) = \sum_{\mathbf{h}} \sum_{\tilde{\mathbf{x}}} P(\mathbf{t}, \mathbf{h}, \tilde{\mathbf{x}} | \theta) P(\tilde{\mathbf{x}} | \mathbf{x}, \mathbf{b}^{(0)}, \mathbf{J}) \quad (6)$$

となる. また, 学習における対数尤度関数は

$$\phi_D(\theta) = \frac{1}{N} \sum_{\mu=1}^N \ln \left(\sum_{\mathbf{h}} \sum_{\tilde{\mathbf{x}}} P(\mathbf{t}^{(\mu)}, \mathbf{h}, \tilde{\mathbf{x}} | \theta) P(\tilde{\mathbf{x}} | \mathbf{x}^{(\mu)}, \mathbf{b}^{(0)}, \mathbf{J}) \right) \quad (7)$$

となる. ここで式 (7) 中の $\tilde{\mathbf{x}}$ に関する期待値は解析的に計算することができないため, モンテカルロ積分による近似計算によって求める. モンテカルロ積分とは, 目的とする関数のある確率分布下での期待値を考える際に, 確率分布から独立に生成したサンプル集合を用いた標本平均によって近似する方法である.

3. 認識実験

MNIST と呼ばれる手書き数字のデータセットを提案法の認識実験を行う. MNIST は本来 60000 個の訓練データがあるが, 本実験では問題設定を難しくするという観点から訓練データを 6000 個に減らして学習を行い, 10000 個のテストデータに対して認識実験を行う.

各モデルの素子数の設定として, 入力層の素子数である $n = 784$, 中間層の素子数である $m = 500$, 出力層の素子数である $k = 10$ とし, PELM の素子数である $\tilde{m} = 500$ とした. また, PELM の近似におけるモンテカルロ積分のサンプル数は 5 としている. 入力データは前処理として標準化を行い, 学習方法として Adam[6] と確率的勾配法を用いている. 確率的勾配法におけるミニバッチサイズは 100 とした. さらに, 学習するパラメータの初期値は Xavier の初期値 [7] で与えた.

ここで, PELM における非学習型パラメータである, バイアスパラメータ $\mathbf{b}^{(0)}$ と結合重みパラメータ $\mathbf{J}$ の決定方法を考える. 本稿においては, 2 種類の手法を用いて性能を比較する. 一つ目はバイアスパラメータをすべて 0, 重みパラメータ $\mathbf{J}$ を

$$G(\mathbf{J} | 0, \sigma^2) = \prod_{i=1}^n \prod_{r=1}^{\tilde{m}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{J_{r,i}^2}{2\sigma^2} \right), \quad \sigma = \frac{1}{\sqrt{n}} \quad (8)$$

から生成されるガウス乱数を用いてランダムに決める方法である. 二つ目は, Gaussian-Bernoulli RBM(GBRBM)[4] というモデルを用いて訓練データを学習し, その GBRBM の隠れ層におけるバイアスパラメータと結合重みパラメータを $\mathbf{b}^{(0)}$ と $\mathbf{J}$ に当てはめるという方法である. GBRBM は可視層と隠れ層の 2 層構造を持ち, 可視層は連続値, 隠れ層が 2 値で与えられるモデルであり, そのためしばしば自己符号化器として利用されている.

なお, GBRBM の学習において生じる厳密計算が難しい期待値については, Contrastive Divergence 法 [5] を用いて近似計算を行った.

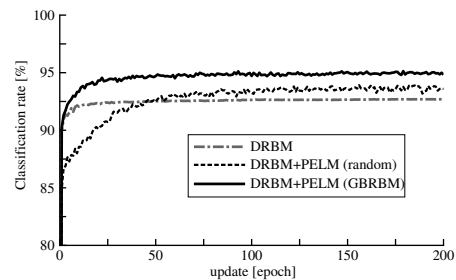


図 4 テストデータにおける認識率の推移.

図 4 に認識実験の結果を示す. グラフにおける横軸が確率勾配法における更新回数 (1 エポック) であり, 縦軸がテストデータに対する認識率である. また, グラフ中の呼称として, 通常の DRBM を "DRBM", ランダムパラメータを用いた提案法を "DRBM+PELM(random)", GBRBM からパラメータを取得した提案法を "DRBM+PELM(GBRBM)" としている. グラフより, 提案法は通常の DRBM よりも最終的な認識精度が高くなる結果となった. また, ランダムパラメータを用いた場合は学習が緩やかに行われるのに対し, GBRBM からパラメータを取得した場合, 学習速度が早くなり, 認識精度が向上した. つまり, 非学習型のパラメータの決定方法がモデルの性能に大きく寄与しているといえる.

4. まとめ

本稿では PELM を用いて DRBM の多層化を行い, 認識実験を行った. その結果, 提案法は多層化前と比較して認識精度が向上することが示された. また, 非学習型パラメータの決定方法が学習性能に大きな影響を及ぼすことも確認された. よって今後の課題として, より良質なパラメータの決定方法を模索することがあげられる.

文献

- [1] 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: 深層学習, 近代科学社, 2015
- [2] H. Larochelle and Y. Bengio: Classification using discriminative restricted Boltzmann machines, Proceedings of the 25th International Conference on Machine Learning, 2008.
- [3] Guang-Bin Huang, Qin-Yu Zhu and Chee-Kheong Siew: Extreme Learning Machine: Theory and applications, Neurocomputing, vol.70, 2006.
- [4] K. Cho, A. Ilin, and T. Raiko: Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines, ICANN2011, 2011
- [5] Hinton, Geoffrey E: Training products of experts by minimizing contrastive divergence, Neural computation, 2002
- [6] Diederik P. Kingma and Jimmy Ba: Adam: A Method for Stochastic Optimization. ICLR, 2015
- [7] Glorot, Xavier, and Yoshua Bengio: Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010.