

# 多層 Bayesian Net の構築に向けた手書き数字からのパターン抽出

Pattern Extraction from handwritten digits images for multi-layer Bayesian Network

西野 兼治      稲葉 真理  
Kaneharu Nishino      Mary Inaba

東京大学大学院 情報理工学系研究科

Graduate School of Information Science and Technology, The University of Tokyo

Brains have hierarchical architectures and information in them flows not only from lower areas to higher areas but also from higher to lower. As the model for this bottom-up and top-down flows, some researchers proposed the hypothesis that information processing in cerebral cortex is based on Bayesian Networks. In this work, we developed the Bayesian AutoEncoder (BAE), which is a method to construct a multi-layer restricted Bayesian Network with hidden variables by extracting patterns from a training dataset. Networks constructed by BAE have hidden variables that represent patterns of the data and can execute inferences about whether the patterns exist or not in each input datum. BAE can construct Bayesian Networks with one hidden parent variables layer from small artificial images and extract its patterns as hidden parent variables. In this paper, we applied BAE to handwritten digit images and confirmed it can extract patterns from natural images.

## 1. はじめに

近年機械学習の分野で深層学習が広く用いられている [Le 13, Hinton 06a, Bengio 07]. 深層学習は入力データ内の特徴やパターンを抽出し, 低い層では局所的かつ単純なパターンを, 高い層ではそれらを組み合わせた複雑なパターンを表すノードを持ち, これらを用いることで高い認識性能を持つ.

このような階層構造は大脳新皮質の視覚野と類似性が指摘されるが, その一方で, 深層学習の多くのニューラルネットは入力層から出力層に一方向に接続されるのに対して, 脳では高次の領野から低次の領野に投射する接続の存在が知られる. これらは注意や予測として認識に影響するとされ, これらトップダウン経路を含めた情報処理モデルとして Bayesian Net を用いる研究がなされている [Doya 07, Rao 05].

これを踏まえ, 我々はパターン認識器として Bayesian Net を採用し, 特徴抽出を行い多層のネットワークを学習する手法 Bayesian AutoEncoder (BAE)[Nishino 15] を開発した. これは訓練データから抽出された各パターンに対応する隠れ変数を持ち, 深層学習と同様に下位層では単純な, 上位層では複雑なパターンを表す変数を得ることを目指している. 構築されるネットワークは確率伝播法による推論を行い, 各パターンが入力データに存在するかを推論し, パターンの認識を行える. この推論はトップダウンな情報とボトムアップな情報を確率論に従って統合して行われるので, 環境や文脈などの大域的な情報を個々の認識に反映でき, より柔軟な認識を期待できる.

これまで人工的に生成した 5x5 ピクセルの画像群を訓練データとして Bayesian Net を構築できた. 本稿では実用的なタスクとして数字認識を設定し, 数字を認識できる Bayesian Net の構築を目指す. その第一段階として手書き数字の画像群からの特徴抽出を行い, BAE の特徴抽出性能を確認する.

## 2. Bayesian AutoEncoder

Bayesian AutoEncoder は訓練データから特徴抽出を行い制限付き 2 値多層 Bayesian Net を構築する手法である. これ

は訓練データに含まれるパターンを表す 2 値変数を隠れ変数に持つネットワークを構築し, 構築されたネットワーク (以下 BAE ネットワーク) は入力データに各パターンが存在するかを確率論に沿った形で推論する.

BAE ネットワークは階層性を持ち, 上位の層は下位層の親変数層となるよう制限される. 最下層を入力層として用い, 他の層は抽出したパターンに対応する隠れ変数の層となる. 隠れ変数の 2 値は対応するパターンが存在する (T), 存在しない (F) を示す. 最下層は入力次元と同じ数の 2 値変数からなり, 対応する入力の値からメッセージを計算する. BAE ネットワークは確率伝播法に基づいて全メッセージを更新し各変数の事後確率を計算し, パターンの有無を認識する.

BAE は教師なし学習によりネットワークを構築する. 学習するパラメータは条件付き確率 (CP) とリンク強度 (LS) の 2 種類である. CP は変数間の条件付き確率を示す. 通常の Bayesian Net では  $n$  個の親を持つ子変数一つにつき  $O(2^n)$  個のパラメータを要するので, 本研究では One-parent-T 仮定と呼ぶ仮定を設けた. これは一つの子を共有する親変数間の高タワツが T を取るとする制限であり, 複数の親が T を取る CP を扱わないことでパラメータ数を  $O(n)$  とした. 同時に, 子を共有する複数の親が T とならないよう推論を限定し認識にスパース性を与える. LS はネットワーク構造の学習のために導入したパラメータであり, 各リンクの必要度を示す. LS の大きなリンクを通るメッセージほど推論に高い影響を持ち, 小さな LS のリンクは影響を与えないよう, メッセージの計算式を変更した. これによりネットワーク構造を離散最適化によらず学習する.

BAE は多層ネットワークの構築の際に深層学習で提案された Greedy Layerwise training[Hinton 06b] に基づいて学習する. この過程を図 1 に示す. 最初は親変数層 1 層と子変数層 1 層の間を全結合した状態から学習をはじめ, 学習が完了したのちに新たな層を上層として追加し, その層について学習を行う. これを繰り返し多層ネットワークを得る.

学習は CP, LS それぞれについて, 各変数ノードで分散して行われる. CP の学習は上下の推論を一致させることを原理として学習を行う. すなわち各変数において, その上位からのメッセージによって計算される事後確率とその下位からのメッセージによって計算される事後確率を一致させるよう

連絡先: 西野 兼治, 東京大学 大学院情報理工学系研究科 創造情報学専攻, nishino.kaneharu@ci.i.u-tokyo.ac.jp

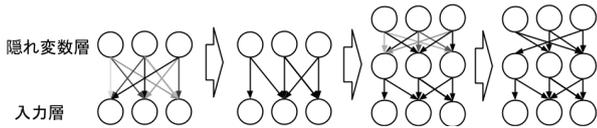


図 1: 多層ネットワークの構築過程. 最下層を入力層として隠れ層のネットワーク構造を学習していく.

に, 上位層との間の CP を更新する. LS は情報量最大化原理 [Linsker 88] によって学習を行う. 学習する最上層の親変数群を出力として, 親変数群と入力との相互情報量を最大化することで学習を行う.

これまでに 5x 5 ピクセルの人工画像からのパターン抽出に隠れ層 1 層で成功し, 親変数層を一層持つ BAE ネットワークの構築が可能である [Nishino 15, 西野 17].

### 3. 手書き数字データセットへの適用

本稿では BAE の実用的なタスクとして文字認識を設定し, その実現のための第一段階として BAE が数字画像群から特徴抽出を行えることを確認する. 入力には手書き数字データセット MNIST \*1 の画像から中央の 24x24 ピクセルの部分抜き出し 12x12 ピクセルに解像度を下げて作成した画像群を用いた. ネットワーク構造は隠れ変数 200 個で学習を行い, 入力層 144 個との間のパラメータについて学習を行った.

パラメータの更新を 180 万回行った時点での学習結果を図 2 に示す. それぞれの枠一つ一つが親変数に対応しており, それぞれの枠内が親変数の持つリンクの状態を示している. それぞれのピクセルは入力画像のそれぞれの位置に対するリンクに対応しており, その色がそのリンクに関するパラメータを示す. ピクセルの赤の値は親変数が T となる条件下でそのピクセル対応する子変数が T を取る条件付き確率を示し, 青はその親変数が T となる条件下で子変数が F となる条件付き確率を示す. また濃度は LS を示し, 黒は LS が低くリンクの存在しない部分, 色の濃い部分は LS が大きくリンクの存続した部分を示す. すなわち赤は興奮性, 青は抑制性の接続と解釈でき, 輝度が接続の強度とみなすことができる.

図 2 に示されるように, 複数の直線状のパターンが親変数として抽出された. MNIST のような手書き数字画像に対しても BAE は適用でき, またこれまでの入力次元 5x5 ピクセル, 隠れ変数 20 個程度の小さなネットワークよりも大規模なネットワークに対しても動作し, パターンを抽出することができることが確認された. その一方で, 得られた親変数は同様のパターンを表現するものが複数存在することが示され, スパース性が得られなかった. LS を 0.5 を閾値として 2 値化した際にリンクの状態が同一となる親変数の数は, 親変数 0 番と同じパターンを持つものが 37 個, 3 番と同じものが 36 個存在する. 2 値化してなお異なるリンクの状態となる親変数の種類は 63 種類であった. 以前に行った 5x5 ピクセルの人工画像を用いた実験 [西野 17] では同一のパターンを示す冗長な親変数は得られずコンパクトなネットワークが得られたが, 今回の結果は冗長性の高い結果となり, スパース性が確保された結果とはならなかった.

また, 学習過程における親変数のリンクの変遷の例を図 3 に示す. 図 3 はそれぞれ 184 番と 135 番の学習途上の変遷で

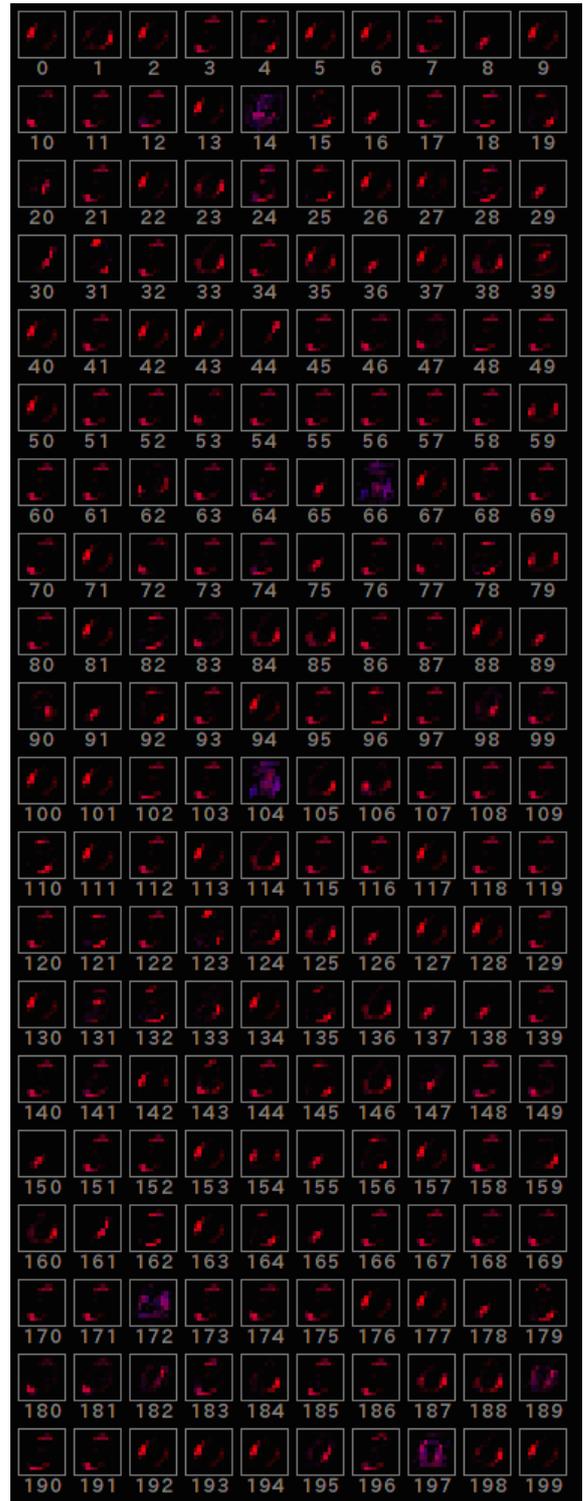


図 2: MNIST から作成した画像群に BAE を適用した結果. それぞれの枠が親変数に対応し, 枠下の番号は親変数の番号を表す. 枠内のピクセルは 12x12 ピクセルの各位置に対するリンクの状態を示す. 黒は接続の無い部分, 赤は CP が高い部分, 青は CP が低い部分を示す.

\*1 MNIST handwritten digit database  
Yann LeCun, Corina Cortes and Chris Burges  
<http://yann.lecun.com/exdb/mnist/>



図 3: 学習におけるリンクの変遷.

あり、更新を 30 万回した時点から 180 万回時点まで 30 万回置きにリンクの状態を示した。135 番は大きな変化をせずにパターンを維持しているが、184 番は大きく変化し、一度 135 番と同様のパターンとなってからまた異なるパターンとなっている。同じパターンとなったのちに変化していることから一定のスパース性は存在するものと思われるが、一度同じパターンに収束することから十分なものではないと考えられる。

スパース性が十分に確保されないことから、One-parent-T 仮定による親変数間の相互抑制性が十分に機能していないことが示唆される。このことは入力を与えた際と同じパターンを持つ親変数間の事後確率からも確認され、それらの合計値が 1.0 を大きく越えることから、一つの親だけが T を取りうるように推論が限定されていないことが明らかとなった。BAE はスパース性の保証を One-parent-T 仮定に相互抑制性に拠るので、十分な相互抑制性を確保することが必要と考えられる。

#### 4. 終わりに

本稿では BAE を MNIST 手書き数字データセットから作成した画像群に対して適用し、パターン抽出を試みた。BAE は直線上のパターンを抽出し部分的にはパターン抽出が行えたと考えられるが、複数の親変数が同一のパターンを表すなど冗長に親変数が獲得された。また一般に基底として知られるガボールフィルタ状のパターンを得ることはできなかった。これらは隠れ変数のスパース性の欠如が原因と考えられる。BAE は One-parent-T 仮定による親変数間の相互抑制によりスパース性を得ているが、今回の実験ではこの相互抑制が十分に機能していないことが判明した。この相互抑制を今後の課題として開発を続けたい。

#### 謝辞

本研究は、総合科学技術・イノベーション会議により制度設計された革新的研究開発推進プログラム (ImPACT) により、科学技術振興機構を通して委託されたものです。

#### 参考文献

- [Bengio 07] Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H.: Greedy layer-wise training of deep networks, in Bernhard Schölkopf, J. P. and Hoffman, T. eds., *Advances in Neural Information Processing Systems 19 (NIPS' 06)*, pp. 153 – 160, MIT Press (2007)
- [Doya 07] Doya, K., Ishii, S., Pouget, A., and Rao, R. P.: *Bayesian brain : probabilistic approaches to neural coding*, Computational neuroscience, Cambridge, Mass. MIT Press (2007)

- [Hinton 06a] Hinton, G. E., Osindero, S., and Teh, Y.-W.: A Fast Learning Algorithm for Deep Belief Nets, *Neural Computation*, Vol. 18, No. 7, pp. 1527–1554 (2006)
- [Hinton 06b] Hinton, G. and Salakhutdinov, R.: Reducing the Dimensionality of Data with Neural Networks, *Science*, Vol. 313, No. 5786, pp. 504 – 507 (2006)
- [Le 13] Le, Q.: Building high-level features using large scale unsupervised learning, in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*, pp. 8595–8598 (2013)
- [Linsker 88] Linsker, R.: Self-organization in a perceptual network, *Computer*, Vol. 21, No. 3, pp. 105–117 (1988)
- [Nishino 15] Nishino, K. and Inaba, M.: Feature Extraction Based on Generating Bayesian Network, in Arik, S., Huang, T., Lai, W. K., and Liu, Q. eds., *Neural Information Processing - 22nd International Conference, ICONIP 2015, Istanbul, Turkey, November 9-12, 2015, Proceedings, Part IV*, Vol. 9492 of *Lecture Notes in Computer Science*, pp. 255–262, Springer (2015)
- [Rao 05] Rao, R. P. N.: Bayesian inference and attentional modulation in the visual cortex, *Neuroreport*, Vol. 16, No. 16, pp. 1843–1848 (2005)
- [西野 17] 西野 兼治, 稲葉 真理: 特徴抽出にむけた情報量最大化原理による Bayesian Network の構造学習, 電子情報通信学会技術研究報告 = IEICE technical report : 信学技報, Vol. 116, No. 424, pp. 71–76 (2017)