

# スパース符号化により画像の分散表現を獲得する 2層ベイジアンネットワーク

Two-layered Bayesian Network Learning Sparse Representation of Image Data

一杉裕志\*1  
Yuuji Ichisugi

\*1産業技術総合研究所 人工知能研究センター  
National Institute of Advanced Industrial Science and Technology(AIST) AIRC

Cerebral cortices might be some kind of restricted, scalable Bayesian networks. In this paper, we examine noisy-OR networks as candidates of scalable Bayesian-network model of cerebral cortex. We have implemented loopy belief propagation algorithms and a gradient-descent learning algorithm for this model.

## 1. はじめに

大脳皮質のすべての領野は似通った解剖学的構造をしていることから、共通の動作原理で動いていると思われる。その共通原理を解明する有力な手がかりとして「大脳皮質がベイジアンネットワークである」という仮説がある [5][7]。我々はこの仮説にもとづいて、ベイジアンネットワークを用いて脳が持つ種々の認知機能を実装し、実用化への道筋を付けることを目指している。

これまで我々は、条件付確率表のモデルに線形和の制約を入れることで推論・学習アルゴリズムを最適化し、ディープラーニングと同じ構造のネットワークのもとで手書き数字認識が行えることを示した [8][9]。しかし認識精度は十分ではなく、その理由の1つとしては、線形和では制約が強すぎて表現力が弱いということが考えられる。

そこで我々は、線形和とは違う制約にもとづく別のスケラブルな条件付確率表の模索をはじめた。その第一歩として今回は noisy-OR モデル [1] を検討した。

## 2. 隠れ変数を含むベイジアンネットワークにおける勾配法

隠れ変数を含むベイジアンネットワークのパラメタ学習は、勾配法により、局所的な情報だけを用いて実行することができる [3]。すべてのパラメタのベクトルを  $\theta$ 、ノード  $X$  の条件付確率表  $P_\theta(x|pa(x))$  を決めるパラメタの1つを  $w$  とし、入力（観測変数に与えられる観測値）のベクトルを  $\mathbf{e}$  とすると、勾配は下記ようになる。

$$\Delta w = \frac{\partial P(\mathbf{e}|\theta)}{\partial w} = \sum_{x, pa(x)} \frac{P_\theta(x, pa(x)|\mathbf{e})}{P_\theta(x|pa(x))} \frac{\partial P_\theta(x|pa(x))}{\partial w} \quad (1)$$

ここで必要となる  $P_\theta(x, pa(x)|\mathbf{e})$  の値は、様々な推論アルゴリズムを用いて効率的に厳密計算または近似計算することができる。

## 3. Noisy-OR モデル

ベイジアンネットワークにおいて各ノードの条件付確率表を素朴に表現すると、親ノードの数  $m$  に対して  $O(2^m)$  個のパラメタが必要になる。これは計算量・メモリ量の爆発や、過適合・

連絡先: 一杉裕志、茨城県つくば市梅園 1-1-1 中央第1  
産業技術総合研究所、y-ichisugi@aist.go.jp

局所解の原因になる。そこで、条件付確率表をより少ないパラメタで表現する場合がある。そのやり方の1つとして下記の式で定義される noisy-OR モデル [1] がある。

$$P_\theta(X=0|U_1=u_1, \dots, U_m=u_m) = \prod_{k=1}^m (1-w_k)^{u_k}$$

$$P_\theta(X=1|U_1=u_1, \dots, U_m=u_m) = 1 - \prod_{k=1}^m (1-w_k)^{u_k} \quad (2)$$

ただし、 $U_k \in \{0, 1\}$  とする。

Noisy-OR は生成モデルとしての表現能力は非常に制限されているものの、「原因のどれか1つが成り立っていれば結果が成り立つ」という自然な関係を簡潔に表現でき、ネットワークの意味が解釈しやすいという利点がある。

Noisy-OR のもう1つの大きな特徴として、スケラビリティがある。ある noisy-OR ノードが親ノードを  $m$  個持つとき、それと等価でノード数  $O(m)$  の二分木のベイジアンネットワークに変換でき、それにより推論・学習アルゴリズムを高速化できる [2]。この方法により、ネットワーク中のエッジの数が  $n$  のとき、多くの推論・学習アルゴリズムの1ステップが  $O(n)$  の計算量で実行できるようになる。

## 4. 実験・評価

### 4.1 推論アルゴリズムの実装と評価

ベイジアンネットワーク上での推論を、局所的な演算の反復により高速に行えるアルゴリズムとして、確率伝搬アルゴリズム [1] がある。Noisy-OR と等価な二分木のベイジアンネットワーク上でのメッセージ計算式を導出し、それに基づいてメッセージ伝播の1ステップを  $O(n)$  ( $n$  はネットワーク中のエッジの数) で実行できる確率伝搬アルゴリズムを実装した。また、比較のために、すでに実装済みの  $O(2^n)$  の素朴な確率伝搬アルゴリズムを nosisy-OR モデルで動作できるようにした。

この2つの推論アルゴリズムの実装を、層構造を持ち、ランダムな層間結合（ループあり）と結合パラメタを持つネットワーク上で、MPE（事後確率が最大となる隠れ変数の値の組）の計算の精度を厳密解と比較する方法で評価した。

その結果、エッジ数の少ない比較的単純なネットワークでは厳密解と一致した推論が行える場合もあったが、ある程度複雑な形のネットワークでは振動が起きやすく、精度が非常に悪いという現象が、2つの実装でともに観察された。

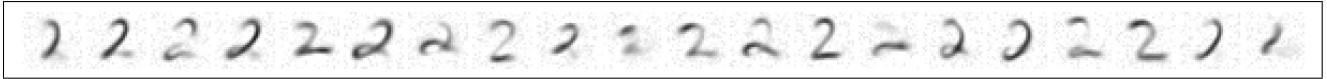


図 1: 手書き数字の「2」の学習で得られた基底画像。平均3個の基底画像の組み合わせで入力画像を近似するように学習。すべての基底画像は独立かつ等頻度で選ばれる。数字を構成するパーツが基底画像として獲得されている。

## 4.2 学習アルゴリズムの実装と評価

ネットワークに入力が与えられた後、推論アルゴリズムにより MPE が求められた時の確率的勾配降下法の学習則を以下に示す。式 (1) に式 (2) を代入し、 $P_{\theta}(x, pa(x)|e)$  の値は  $x, pa(x)$  の値の組が MPE と一致するときのみ 1 で他の場合は 0 であるとみなすと、下記の学習則が導かれる。

$$\Delta w_k = \begin{cases} 0 & (u_k = 0) \\ \frac{-\prod_{i \neq k} (1-w_i)^{u_i}}{\prod_i (1-w_i)^{u_i}} & (x = 0, u_k = 1) \\ \frac{\prod_{i \neq k} (1-w_i)^{u_i}}{1 - \prod_i (1-w_i)^{u_i}} & (x = 1, u_k = 1) \end{cases} \quad (3)$$

この学習則を用いて、2層 noisy-OR ネットワークによる MNIST 手書き数字の教師なし学習を行った。

MPE の推論アルゴリズムには山登り法 [6] を使用した。正則化機構としては、勝率ペナルティと側抑制ペナルティ [8] を用いた。

図 1 は、隠れ層のノード数 20 個で、入力層の 28x28 個の 2 値ノードに手書き数字の「2」のピクセル値を与え、隠れ層の活性ノード数を平均3個としたときの、条件付確率表の学習結果である。基底画像（親ノード）とピクセル値（子ノード）の間の noisy-OR の各パラメタの値を、濃淡として可視化した。パーツを表現する基底画像の組み合わせで入力を表現するように学習が進んでいることが見て取れる。

## 5. 関連研究

2層の noisy-OR ネットワークにおける教師なし学習は以前から行われているが、先行研究では大規模オンライン学習は必ずしも指向していない。

NMF(非負値行列因子分解)[4] は、入力を非負の基底ベクトルの非負の重み付き線形和で近似するという点で、今回の実験に用いたネットワークの振る舞いと定性的に似ている。しかし、ベイジアンネットを使う方法は、noisy-OR 以外の条件付確率表モデルを持ったノードと組み合わせたり、多層化や、他の様々な正則化の機構と組み合わせることが容易であり、拡張性が高いと言える。

## 6. まとめと今後

大規模ベイジアンネットを用いた様々な認知機能の実現を可能とするために、本稿ではスケーラブルな条件付確率表のモデルの1つである noisy-OR モデルの推論・学習アルゴリズムを実装し、予備的実験により動作確認を行った。ループがある noisy-OR ネットワークのもとでは、特に工夫をしない確率伝搬アルゴリズムは振動し非常に推論の精度が悪いことが観察された。2層 noisy-OR ネットワークにおける確率的勾配降下法を用いた教師なし学習は、適切な正則化を行うことで、入力画像をスパース符号化した表現を獲得できることを確認した。

振動に対しては、メッセージ伝搬のスケジューリングを工夫したり、ダンピングと呼ばれる方法でメッセージの値の変化の平滑化を行うことで軽減できることが知られており、今後の検討が必要である。

Noisy-OR と同様なスケーラビリティを持つ条件付確率表のモデルは noisy-AND など他にも無数に考えられるため、それらの工学的有用性や神経科学的妥当性も今後検討していく。

## 謝辞

産総研の佐野崇氏、川田正晃氏にはベイジアンネットのオンライン学習手法の調査に協力していただいた。また本村陽一氏には noisy-OR の表現力等に関して有益な助言をいただいた。

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものである。

## 参考文献

- [1] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.
- [2] David Heckerman, Causal Independence for Knowledge Acquisition and Inference, In Proc. of UAI-93, 122-127, 1993
- [3] J. Binder, D. Koller, S.J. Russell, and K. Kanazawa Adaptive probabilistic networks with hidden variables, Machine Learning, 29(2-3), 213-244, 1997.
- [4] Daniel D. Lee and H. Sebastian Seung, Learning the parts of objects by non-negative matrix factorization Nature 401, 788-791, 1999.
- [5] T.S. Lee, D. Mumford, Hierarchical Bayesian inference in the visual cortex. Journal of Optical Society of America, A 20(7): pp.1434-1448, 2003.
- [6] 一杉裕志、「大脳皮質のアルゴリズム BESOM Ver.1.0」, 産業技術総合研究所テクニカルレポート AIST09-J00006, Sep 2009.
- [7] 一杉裕志, 解説: 大脳皮質とベイジアンネット、日本ロボット学会誌 Vol.29 No.5, pp.412-415, 2011.
- [8] 一杉裕志、「制限付きベイジアンネット BESOM の正則化の一手法」第96回人工知能基本問題研究会 (SIG-FPAI), 2015.
- [9] Yuuji Ichisugi and Naoto Takahashi, An Efficient Recognition Algorithm for Restricted Bayesian Networks, In Proc. of 2015 International Joint Conference on Neural Networks (IJCNN), 2015.