

Noisy-OR, Noisy-AND ゲートによる位置不変性の変分学習

Variational Learning of Feature Pooling in a Bayesian Network with Noisy-OR and Noisy-AND Gates

佐野崇^{1*} 一杉裕志²
Takashi Sano¹ Yuuji Ichisugi²

¹ 東洋大学 情報連携学部 情報連携学科

¹ Department of Information Networking for Innovation and Design,
Faculty of Information Networking for Innovation and Design,
Toyo University

² 産業技術総合研究所 人工知能研究センター

² National Institute of Advanced Industrial Science and Technology (AIST),
Artificial Intelligence Research Center

Abstract: In the viewpoint of the Bayesian brain hypothesis, Bayesian network model of cerebral cortex is promising not only for computational modeling of brain, but also for an efficient brain-like artificial intelligence. A notorious drawback in a Bayesian network is, however, the number of parameters that grows exponentially against the number of parent variables for a random variable. Restriction of the model may be a solution to this problem. Inspired by the biological plausibility, we previously proposed to use the combination of the noisy-OR and noisy-AND gates, whose numbers of parameters grow linearly with the number of parent random variables. Although we showed that this model can have translation invariance in a small-scale setting, it was difficult to enlarge the scale because of the hidden variables. In this study, we extend the previous attempt by employing a variational learning method to overcome the intractability of the estimation of the massive hidden variables. We can scale the model up to learn the hand-written digit data.

1 はじめに

人間の hochi 機能を司る大脳皮質の計算原理の解明は、汎用人工知能実現の近道である。大脳皮質はベイズ的な確率情報処理を担うと考えられている [1, 2, 3] ため、グラフ上の局所演算として確率的推論を行うベイジアンネットワーク [4] を用いたモデル化が有力であり、脳型人工知能としても有望である [1, 5, 6]。

一般に、隠れ変数を多く含むベイジアンネットワークは、推論や学習が困難である。一方で、脳では、ベイジアンネットワークに何らかの制限があるために、効率的な推論や学習が可能になっていると我々は考えている。大脳の構造を模倣した制限付きベイジアンネットワークとして、著者らはこれまで、BESOM(Bidirectional

Self-Organizing Map) という、大脳の構造を模倣した制限付きベイジアンネットワークを開発してきた [5, 6]。

BESOM の現在のバージョンでは、神経細胞の樹状突起における演算 [7] の考察から、noisy-OR と noisy-AND という 2 種類の条件付き確率表 (CPT) を組み合わせて用いる [8, 9]。Noisy-OR、noisy-AND とともに、1 つの CPT のパラメータ数が、親変数の数について線形にしか増大せず、大規模化に向けた CPT モデルである。ここでは、noisy-OR だけでは表現力が不十分であるが、noisy-AND と組み合わせることで表現力が向上すると考えており、[8, 9] では、このモデルによって構文解析の問題が扱えることが確かめられた。

我々は以前の研究 [10] において、BESOM によって視覚における特徴プーリング機構を再現可能であることを、小規模なモデルを勾配学習することによって確かめた。しかし、このモデルには隠れ変数が含まれているため、大規模化した場合には単純な勾配学習が難しくな

* 連絡先： 東洋大学情報連携学部
115-0055 東京都北区赤羽台
E-mail: takashi.sano@iniad.org

る。そこで本研究では、変分法に基づく近似学習法を用いた学習法を提案する。変分学習法によって隠れ変数についての和が不要になり、大規模化が可能になる。実際に、この方法によって、手書き文字の学習が可能にほど大規模なモデルの学習が可能になった。

2 モデル

2.1 Noisy-OR, Noisy-AND ゲート

確率変数 $X \in \{0, 1\}$, $\mathbf{Z} = \{Z_1, \dots, Z_N\} \in \{0, 1\}^N$ に対し、noisy-OR ゲートは次の形の条件付き確率分布である [4]。

$$P_{\text{OR}}(X = 1|\mathbf{Z}) = 1 - (1 - \mu_0) \prod_{i=1}^N (1 - \mu_i)^{Z_i} \quad (1)$$

$$P_{\text{OR}}(X = 0|\mathbf{Z}) = 1 - P_{\text{OR}}(X = 1|\mathbf{Z}) \quad (2)$$

パラメータ $\mu_i \in [0, 1] (i = 1, \dots, N)$ は親変数 Z_i の寄与の大きさを表す。 μ_0 はリーク確率を表し、親変数の状態に関わらず X が活性化する確率に対応する。

同様に、Noisy-AND ゲートは次のように定義される条件付き確率分布である [4, 11]。

$$P_{\text{AND}}(X = 1|\mathbf{Z}) = (1 - \nu_0) \prod_{i=1}^N (1 - \nu_i)^{1-Z_i} \quad (3)$$

$$P_{\text{AND}}(X = 0|\mathbf{Z}) = 1 - P_{\text{AND}}(X = 1|\mathbf{Z}) \quad (4)$$

$\nu_i \in [0, 1]$ はパラメータであり、noisy-OR ゲートと同様に親変数の寄与の大きさを表す。Noisy-OR ゲートとは、 $X = 0, 1$ の確率と \mathbf{Z} の真偽値が反転した関係になっている。この双対関係は、論理 OR と論理 AND の間のド・モルガンの法則と対応している。

Noisy-OR, noisy-AND ゲートはどちらも、親変数の個数 N に対して $N + 1$ 個のパラメータを持つ。これは、完全な CPT が $\mathcal{O}(2^N)$ 個のパラメータを必要とすることと対照的である。さらに、 $N + 1$ 個のパラメータの解釈も容易である。すなわち、 N 個のパラメータはそれぞれ、 N 個の親変数が子変数の状態決定にどの程度寄与するかを表している。

2.2 提案アーキテクチャ

本研究では、noisy-OR と noisy-AND を組み合わせた、次のような階層的なベイジアンネットワークを用い

る [10]。

$$P(\mathbf{X}, \mathbf{U}, \mathbf{C}, \mathbf{G}) = \prod_i P_{\text{OR}}(X_i|\mathbf{G}) \times \prod_j P_{\text{AND}}(G_j|\mathbf{U}, \mathbf{C}) \times P(\mathbf{C})P(\mathbf{U}) \quad (5)$$

$P(\mathbf{C}), P(\mathbf{U})$ は事前分布である。このモデルの構造を図 1 に示した。

中間層の隠れ変数 \mathbf{G} は、入力画像 \mathbf{X} の特徴を学習する。最上位の変数には、 \mathbf{U} と \mathbf{C} の 2 種類がある。 \mathbf{U} は、入力 \mathbf{X} の高次の特徴 (例えば \mathbf{X} が手書き文字であれば文字種など) を表す変数であり、 \mathbf{C} はその位置を表す変数である。パラメータがうまく学習されれば、変数 \mathbf{U} と \mathbf{C} の組み合わせ (AND 演算) によって \mathbf{G} が生成され、そこから対応する画像の重ね合わせ (OR 演算) によって、可視層 \mathbf{X} に画像が生成される。また、認識時には、 \mathbf{U} は入力 \mathbf{X} に含まれる特徴に対し位置不変に応答し、 \mathbf{C} は位置にのみ依存して応答するため、特徴のプーリングと同じ機能を持つと期待できる。

2.3 学習

本研究で用いるモデルは多数の隠れ変数を含むため、単純な尤度最大化による学習は難しい。そこで、Neural Variational Inference Learning (NVIL)[12] と呼ばれる手法を用いて近似的な学習を行う [13]。NVIL では、生成モデル $P(\mathbf{X}, \mathbf{Z}) = P(\mathbf{X}|\mathbf{Z})P(\mathbf{Z})$ の学習の際に、隠れ変数を推定する変分分布 $Q(\mathbf{Z}|\mathbf{X})$ を導入する。このとき、変分分布 $Q(\mathbf{Z}|\mathbf{X})$ は、 \mathbf{X} を入力とするニューラルネットワークとしてモデル化され、 $Q(\mathbf{Z}|\mathbf{X})$ の学習はニューラルネットワークの学習に帰着される。真の目的関数である対数尤度に対して、変分法から導かれる下限を近似目的関数として、 $P(\mathbf{X}, \mathbf{Z})$ と $Q(\mathbf{Z}|\mathbf{X})$ は同時に最適化される。

提案モデルの実際の学習は、noisy-OR ゲートの学習と、noisy-AND ゲートの学習の 2 段階に分けられる: まず、可視変数 \mathbf{X} を学習データとし、noisy-OR ゲートの重みを NVIL によって教師なしで学習する。このとき、同時に変分分布 $Q(\mathbf{G}|\mathbf{X})$ も学習される。次に noisy-AND ゲートの学習を行う。ここでは、学習データとして、可視変数 \mathbf{X} と、noisy-AND ゲートの親変数 \mathbf{U}, \mathbf{C} の値を用いる。与えられた \mathbf{X} から、学習済みの $Q(\mathbf{G}|\mathbf{X})$ を用いて、隠れ変数 \mathbf{G} の分布を推定する。同時に、 \mathbf{U}, \mathbf{C} を用いて、確率分布 $P_{\text{AND}}(\mathbf{G}|\mathbf{U}, \mathbf{C})$ を計算する。この 2 つの \mathbf{G} の確率分布が一致するよう、交差エントロピーを目的関数として、noisy-AND ゲートのパラメータを

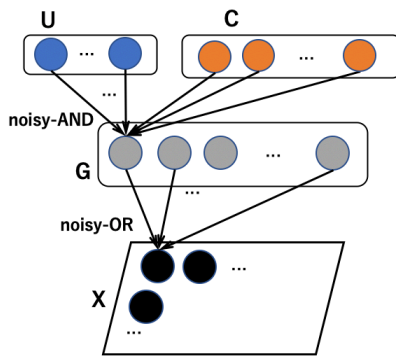


図1 位置不変な特徴をプーリングするベイジアンネットワーク

最適化する。

3 関連研究

NVIL法は、sigmoid belief network というベイジアンネットワークの学習のために開発された [12]。Sigmoid belief network も noisy-OR ゲートや noisy-AND ゲートと同様に、パラメータ数が親変数の数に対して線形にしか増えない点が類似している。しかし、親変数の OR や AND 演算といった解釈ができなため、親変数が子変数に果たす役割を理解することが難しいという違いがある。

変分法に基づく近似によって noisy-OR ゲートの教師なし学習を行う先行研究には [14] がある。この文献では、NVIL法と同様の近似目的関数を最適化することで学習を行っているが、変分分布とモデルのパラメータを交互に更新する変分 EM 法に近いアルゴリズムが用いられている。NVIL法を noisy-OR ゲートの学習に適用した予備的な研究に [13] がある。

Recursive Cortical Network (RCN) はベイジアンネットワークによって脳の視覚系を模倣した高性能なモデルであり、提案モデルとも関連が深い [15]。RCN はスパースな接続構造を持つため、確率伝搬法によって高速な推論が可能なのである。RCN は学習に、尤度最大化とは異なるアドホックなアルゴリズムを用いているため、標準的な尤度最大化法に基づく提案法の方が学習の見通しがよく、拡張性も高いと考えられる。

4 数値実験と結果

ここでは、MNIST 手書き文字 [16] の 60000 枚の学習データを用いた。位置不変性の学習のために、60000 枚の画像をそれぞれ右に 6 ピクセルずらした画像を作成し、合計で 120000 枚の画像を学習に用いた。

ベイジアンネットワークは、 \mathbf{X} を入力に合わせて 784 次元とし、隠れ変数 \mathbf{G} を 32 次元とした。変数 \mathbf{U} は 0



図2 noisy-OR ゲートの獲得した表現の一部

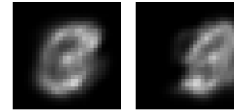


図3 最上位層の変数から生成した画像。2つの画像は変数 \mathbf{C} を文字 0 に固定し、変数 \mathbf{U} をそれぞれ中央と右に変えて生成した。

から 9 までの文字種に対応させ、10 次元とし、変数 \mathbf{C} は、文字が中央にある場合と、右にある場合の 2 通りを表す 2 次元とした。 \mathbf{U} , \mathbf{C} には、教師データを one-hot 表現に変換して与えた。

noisy-OR ゲートの学習では、120000 枚のデータをすべて用い、NVIL によって教師なし学習を行った。学習後、隠れ変数 \mathbf{G} が獲得した表現を、 \mathbf{G} のうち 1 つだけを 1 とし、残りを 0 として可視変数 \mathbf{X} を生成することで可視化した (図 2)。手書き文字に特徴的な線分が学習できていることがわかる。また、画像が中央にある場合と、右にある場合に対応して、線分が中央にある表現と、右側にある表現が存在する。

次に、noisy-AND ゲートの学習を行った。学習には、0 から 9 までの画像を 1 枚ずつ用い、それらを右にシフトさせた画像と合わせて 20 枚を用いた。学習後、学習された表現を、親変数 \mathbf{U} と \mathbf{C} の組み合わせを与え、隠れ変数 \mathbf{G} 、可視変数 \mathbf{X} の順でサンプルし可視化したものを示した (図 3)。生成された画像はあまり鮮明ではないが、同じ特徴 \mathbf{U} に対し、位置変数 \mathbf{C} を中央と右に変えると、生成される画像の位置が変化の様子が確認できる。

5 まとめ

Noisy-OR と noisy-AND ゲートを組み合わせ、視覚系を模した階層型ベイジアンネットワークモデルを提案した。このモデルは多数の隠れ変数を持つため、周辺化を行って尤度最大化による学習を行うことが困難であるが、変分学習法である NVIL 法を用いることで、手書き数字が学習可能な規模にまでモデルをスケールアップすることができた。Noisy-OR ゲートは教師なし学習によって、手書き文字に典型的な線分の特徴を獲得した。Noisy-AND ゲートは、文字種とその位置を教師データとして与えて学習することで、特定の文字を特定の位置に生成することが可能になった。

本研究では画像生成の実験によって位置不変性の獲得

の定性的な確認を行ったが、より正確には画像の認識によって位置不変性の確認をする必要があり、今後の課題である。また、生成された画像もあまり鮮明ではなく、ハイパーパラメータの調整などでより良い学習ができるかどうか検討する必要がある。

本研究ではモデルの表現力を確認するために、変分近似である NVIL 法を用いた学習を行ったが、このような学習機構を脳が採用していると主張するものではない。むしろ NVIL 法は誤差逆伝播法を用いており、生物学的に妥当であるとは言いがたい。より生物学的に妥当な方法は、ベイジアンネットワーク上の局所計算で推論が行えるループあり確率伝搬法 [17] を用いた学習法であり、今後検討していきたい。

謝辞

産総研の高橋直人氏、中田秀基氏、竹内泉氏との議論に深く感謝いたします。

本研究は JSPS 科研費 JP18K18117, JP18K11488 の助成を受けたものです。

参考文献

- [1] T. S. Lee and D. Mumford. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, Vol. 20, No. 7, pp. 1434–1448, 2003.
- [2] D. C. Knill and A. Pouget. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, Vol. 27, No. 12, pp. 712–719, 2004.
- [3] K. Doya, S. Ishii, A. Pouget, and R. P. Rao. *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT press, 2007.
- [4] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2 edition, 1988.
- [5] Y. Ichisugi. A cerebral cortex model that self-organizes conditional probability tables and executes belief propagation. In *IEEE International Conference on Neural Networks - Conference Proceedings*, pp. 178–183, 2007.
- [6] 一杉裕志. 解説:大脳皮質とベイジアンネットワーク. 日本ロボット学会誌, Vol. 29.5, pp. 412–415, 2011.
- [7] C. Koch, T. Poggio, and V. Torre. Retinal ganglion cells: a functional interpretation of dendritic morphology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, Vol. 298, No. 1090, pp. 227–263, 1982.
- [8] 一杉裕志. 疑似ベイジアンネットワークを用いた認知モデルのプロトタイピング手法の提案. 第 4 回 人工知能学会 汎用人工知能研究会 (SIG-AGI), 2016.
- [9] N. Takahashi and Y. Ichisugi. Restricted Quasi Bayesian Networks as a Prototyping Tool. In *The Third Workshop on Advanced Methodologies for Bayesian Networks (AMBN)*, 2017.
- [10] 佐野崇, 一杉裕志. ベイジアンネットワークにおける noisy-or, and ゲートによる特徴のプーリング. 第 9 回 人工知能学会 汎用人工知能研究会 (SIG-AGI), 2018.
- [11] F. J. Diez. Canonical Probabilistic Models for Knowledge Engineering. *Information Sciences*, Vol. 9, pp. 1–59, 2007.
- [12] A. Mnih and K. Gregor. Neural Variational Inference and Learning in Belief Networks. In *the 31st International Conference on Machine Learning (ICML)*, pp. 1792–1799, 2014.
- [13] 佐野崇, 一杉裕志. ニューラルネットワーク最適化による noisy-or ベイジアンネットワークの変分ベイズ学習. 第 29 回 日本神経回路学会 全国大会, 2019.
- [14] T. Šingliar and M. Hauskrecht. Noisy-OR Component Analysis and its Application to Link Analysis. *Journal of Machine Learning Research (JMLR)*, Vol. 7, pp. 2189–2213, 2006.
- [15] D. George, W. Lehrach, K. Kanksy, M. Lázaro-Gredilla, C. Laan, B. Marthi, X. Lou, Z. Meng, Y. Liu, H. Wang, A. Lavin, and D. S. Phoenix. A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs. *Science*, Vol. 358, No. 6368, 2017.
- [16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.
- [17] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy Belief Propagation for Approximate Inference: An Empirical Study. In *UAI'99: Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pp. 467–475, 1999.