Restricted Boltzmann Machine の学習における 重み付けサンプリングと正則化項の効果

D－20

Effectiveness of Weighted Sampling and Regularization for Training Restricted Boltzmann Machines

吉本 昌史† 小林 学‡
Masashi YOSHIMOTO† and Manabu KOBAYASHI‡

†湘南工科大学大学院 工学研究科 電気情報工学専攻 ‡湘南工科大学 工学部 情報工学科
†Graduate School of Engineering, Shonan Institute of Technology ‡Shonan Institute of Technology

1. はじめに

Hinton らによって提案された Contrastive Divergence (CD)法により，Restricted Boltzmann Machine(RBM)が高速に学習可能であることが示され，Deep Learning へ効果的に応用された[1]．またこれらを改良する方式が種々提案されている[2-4]. 本研究では RBM の学習において微分値の近似値を求めるサンプリングに重み付けを行う．また RBM の目的関数に正則化項を導入し，その効果を明らかにする．

2. Restricted Boltzmann Machine(RBM)

データを$\boldsymbol{x} = (x_i)_{i\in V}$，隠れ状態ベクトルを$\boldsymbol{h} = (h_j)_{j\in H}$とし，パラメータ$W = [w_{ij}]$とする．ここで

$$P(\boldsymbol{x},\boldsymbol{h}|W) = \frac{1}{Z}e^{-E(\boldsymbol{x},\boldsymbol{h})} \tag{1}$$

$$E(\boldsymbol{x},\boldsymbol{h}) = \sum_{i\in V}\sum_{j\in H} x_i w_{ij} h_j = -\boldsymbol{x}^{\mathrm{T}} W \boldsymbol{h} \tag{2}$$

$$Z = \sum_{\boldsymbol{x}'}\sum_{\boldsymbol{h}'} e^{-E(\boldsymbol{x}',\boldsymbol{h}')} \tag{3}$$

と定義する．RBM の概略を図 1 に示す．このとき RBM の目的関数を

$$\log P(\boldsymbol{x}|W) = \log \textstyle\sum_{\boldsymbol{h}'} P(\boldsymbol{x},\boldsymbol{h}'|W) = \log \sum_{\boldsymbol{h}'} \frac{1}{Z} e^{-\boldsymbol{E}(\boldsymbol{x},\boldsymbol{h}')}$$
$$= \log \textstyle\sum_{\boldsymbol{h}'} e^{-\boldsymbol{E}(\boldsymbol{x},\boldsymbol{h}')} - \log \sum_{\boldsymbol{x}',\boldsymbol{h}'} e^{-\boldsymbol{E}(\boldsymbol{x}',\boldsymbol{h}')} \tag{4}$$

とする[1-4]．パラメータによる微分は以下となる．

$$\frac{\partial P(\boldsymbol{x},W)}{\partial w_{ij}} = x_i P\left(h_j = 1 \middle| \boldsymbol{x},W\right) - P(x_i' = 1, h_j' = 1|W) \tag{5}$$

このとき次式が成り立つことが知られている．

$$P\left(h_j = 1\middle|\boldsymbol{x},W\right) = \frac{1}{1+e^{-\sum_i w_{ij}x_i}} = g\left(\sum_i w_{ij}x_i\right) \tag{6}$$

ただし$g(x) = \frac{1}{1+e^{-x}}$はシグモイド関数である．ここで学習時において$P(x_i' = 1, h_j' = 1|W)$の項は正確に求めることができない．そこで文献[2-4]ではサンプリングを行うことにより，近似値を求めている．

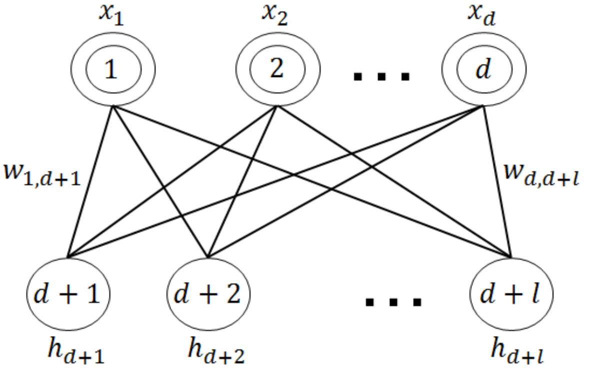

図 1：RBM の概略

3. 重み付けサンプリングと正則化項の導入

本研究では RBM の目的関数を修正して以下とする．

$$\log P(\boldsymbol{x},W) = \log P(\boldsymbol{x}|W) + \log P(W) \tag{7}$$

このとき$-\log P(W)$は正則化項と呼ばれる．良く利用される正則化項としては例えば

$$-\log P(W) = \gamma \sum_{i,j} |w_{ij}| \tag{8}$$

などがある．簡単のため，$D = \left[\frac{\partial \log P(W)}{\partial w_{ij}}\right]$と定義する．

ここで Salakhutdinov らのアルゴリズム[4]を改良する．N個の学習データを$\{x^{(1)},\cdots,x^{(N)}\}$とし，初期パラメータ$W^{(0)}$とする．また初期の$M$個の定常状態サンプルを$(\tilde{x}^{(0,1)},\tilde{h}^{(0,1)}),\cdots,(\tilde{x}^{(0,M)},\tilde{h}^{(0,M)})$としたとき，提案アルゴリズムを以下に示す．

[アルゴリズム]

for($t = 0; t < T$(number of iterations)$; t++$){
  for($n = 1; n \le N; n++$) $\mu_j^{(n)} = g\left(\sum_i w_{ij}^{(t)} x_i^{(n)}\right)$;
  for($m = 1; m \le M; m++$){
    Sample $\left(\tilde{\boldsymbol{x}}^{(t+1,m)}, \tilde{\boldsymbol{h}}^{(t+1,m)}\right)$ given $(\tilde{x}^{(t,m)}, \tilde{\boldsymbol{h}}^{(t,m)})$ by running a Gibbs sampler.
  }
  $W^{(t+1)} = W^{(t)} + \alpha_t\left\{A^{(t+1)} - B^{(t+1)}\right\} + D$
}

ただし

$$A^{(t+1)} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}^{(n)} (\boldsymbol{\mu}^{(n)})^{\mathrm{T}} \tag{9}$$

$$B^{(t+1)} = \frac{\beta_t}{M}\textstyle\sum_{m=1}^{M} \tilde{\boldsymbol{x}}^{(t+1,m)}\left(\tilde{\boldsymbol{h}}^{(t+1,m)}\right)^{\mathrm{T}} + (1-\beta_t)B^{(t)} \tag{10}$$

である．$D = 0, \beta_t = 1$と設定すると従来法と同一となる．

4. まとめ

本研究では RBM の学習において正則化項を導入し，またサンプリングの重み付けを行う手法を提案した．β_tや正則化項のγの問題に適した適切な設定の仕方については，今後の課題である．

参考文献

[1] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” Neural Computation, Vol. 18, 2006.

[2] T. Tieleman, “Training restricted Boltzmann machines using approximations to the likelihood gradient,” In Machine Learning: Proceedings of the Twenty-First International Conference New York: ACM, pp. 1064-1071, 2008.

[3] R. R. Salakhutdinov, and G. E. Hinton, “Deep Boltzmann machines,” In Proceedings of the International Conference on Artificial Intelligence and Statistics, Vol. 12, pp. 448-455, Cambridge, MA: MIT Press, 2009.

[4] R. R. Salakhutdinov, and G. E. Hinton, “An efficient learning procedure for deep boltzmann machines,” Neural Computation, Vol.24, No.8, pp.1967-2006, 2012.