Detailed Balance Learning によるマルコフ連鎖の学習
Detailed Balance Learning for learning of Markov chain

前田新一 (PY)†, 青木佑紀 ‡, 石井信 †
Shin-ichi Maeda (PY), Yuki Aoki, and Shin Ishii
† 京都大学情報学研究科
‡ 京都大学医学研究科
ichi@sys.i.kyoto-u.ac.jp

Abstract— We propose a new learning framework named 'Detailed Balance Learning' (DBL) to learn a stationary distribution of Markov chain, and show DBL has a close relationship with contrastive divergence learning when applied to restricted Boltzmann machine. A sufficient condition for the convergence is also presented.

Keywords— **Markov chain, Detailed balance, Contrastive Divergence, Restricted Botlzmann Machine**

1 序論

既約なマルコフ連鎖は，唯一の定常分布をもつ．しかし，マルコフ連鎖の訪れる状態数が大きいとき，その定常分布を陽に求めることは困難であり，定常分布がサンプルを生成する真の分布に近づくようにマルコフ連鎖のパラメータを学習させることは難しい．本研究では，このようなマルコフ連鎖の定常分布を真の分布に近づけるための新しい学習法として Detailed Balance Learning (DBL) を提案する．ボルツマンマシンなどの離散分布においては，しばしば，同時分布を定義した時の正規化定数をパラメータの関数として陽に表現することが困難になる．このようなモデルにおいて条件付き分布は，例えば，各離散変数が二値であれば二項分布であるように容易に求めることができる．ここで提案する DBL は，このような条件付き分布を用いたギブスサンプリングによるマルコフ連鎖にも適用できるため，ボルツマンマシンなどの正規化定数の計算が困難な離散変数の分布の学習に適用できる．また，この学習法は Contrastive Divergence Learning (CDL) [1] と密接な関係をもつ．

2 Detailed Balance Learning

2.1 コスト関数

状態集合 S を加算集合とし，$\mathbf{v} \in S$ と $r(\mathbf{v})$ をそれぞれ観測変数とその観測変数のしたがう真の分布とする．いま，$\mathbf{v}, \mathbf{v}' \in S$ に関するマルコフ連鎖が，$p(\mathbf{v}|\mathbf{v}', \theta)$ のようにパラメータ θ でパラメトライズされているとすると，ここでの目的は，$p(\mathbf{v}|\mathbf{v}', \theta)$ が唯一の定常分布 $p_\infty(\mathbf{v}|\theta)$ をもつとして，$p_\infty(\mathbf{v}|\theta)$ が真の分布 $r(\mathbf{v})$ に近づくようにパラメータ θ を学習することである．この目的を達成するために，詳細釣り合い条件を用いる．分布 $q(\mathbf{v})$ に対する詳細釣り合い条件は以下で与えられる．

$$\text{For any } \mathbf{v}, \mathbf{v}' \in S,\ p(\mathbf{v}'|\mathbf{v}, \theta) q(\mathbf{v}) = p(\mathbf{v}|\mathbf{v}', \theta) q(\mathbf{v}') \quad (1)$$

もし，上記のような分布 $q(\mathbf{v})$ が存在したならば，$q(\mathbf{v})$ は定常分布 $p_\infty(\mathbf{v}|\theta)$ に一致する．この事実に基づき，以下のコスト関数を考える．

$$F(\theta, \bar{\theta}) = KL\left[p(\mathbf{v}'|\mathbf{v}, \bar{\theta}) r(\mathbf{v}) | p(\mathbf{v}|\mathbf{v}', \theta) r(\mathbf{v}')\right] \quad (2)$$

ここで，$KL[p(\mathbf{x})|q(\mathbf{x})] \equiv \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})}$ は，Kullback-Leibler 擬距離である．同様に，隠れ変数 $\mathbf{h}$ を介したマルコフ連鎖が定義される場合，(狭義の) 詳細釣り合い条件を考えることで，以下のコスト関数を考えることができる．

$$F(\theta, \bar{\theta}) = KL\left[p(\mathbf{v}', \mathbf{h}|\mathbf{v}, \bar{\theta}) r(\mathbf{v}) | p(\mathbf{v}, \mathbf{h}|\mathbf{v}', \theta) r(\mathbf{v}')\right] \quad (3)$$

これらのコスト関数 (2), (3) は，任意のパラメータ θ, $\bar{\theta}$ に対して $F(\theta, \bar{\theta}) \geq 0$ であり，$F(\theta, \theta) = 0$ となるのは，$r(\mathbf{v})$ に対して詳細釣り合い条件が成り立つときのみ，すなわち，少なくとも $r(\mathbf{v}) = p_\infty(\mathbf{v}|\theta)$ が成り立つときのみである．これらの性質から，コスト関数 $F(\theta, \theta)$ が小さくなるようパラメータ θ を学習することに合理性がある．しかしながら，$F(\theta, \theta)$ には，未知の分布 $r(\mathbf{v})$ が含まれるため直接の最小化は難しい．次節で $F(\theta, \theta)$ を最小化するためのアルゴリズムを述べる．

2.2 Detailed Balance Learning (DBL) アルゴリズム

学習において，真の分布 $r(\mathbf{v})$ は未知であるが，$r(\mathbf{v})$ からの独立なサンプルがしばしば利用できる．これらのサンプルを用いることで，$r(\mathbf{v})$ に関する期待値をサンプル平均によって近似することができる．そこで，以下のようなアルゴリズムを考える．

Detailed Balance Learning アルゴリズム

1. $t = 1$ として，θ_0, θ_1 に初期値を設定する．
2. $F(\theta_{t+1}, \theta_t) < F(\theta_{t-1}, \theta_t)$ を満たす θ_{t+1} を求める．
3. 終了条件を満たしたならばアルゴリズムを終了し，そうでないならば t を 1 増やし，ステップ 2 に進む．

$F(\theta, \bar{\theta}) = \sum_{\mathbf{v}'} \sum_{\mathbf{v}} p(\mathbf{v}'|\mathbf{v}, \bar{\theta}) r(\mathbf{v}) \log p(\mathbf{v}|\mathbf{v}', \theta) + const$ (ただし，$const$ は θ に依存しない定数) より，ス

テップ 2 は $\sum_{\mathbf{v}'}\sum_{\mathbf{v}} p(\mathbf{v}'|\mathbf{v},\theta_t)r(\mathbf{v})\log p(\mathbf{v}|\mathbf{v}',\theta_{t+1})$ が評価できれば実行可能であり，これはサンプル平均によって評価できる．また，コスト関数が θ に関して微分可能であれば，ステップ 2 の実行に準ニュートン法などの最適化手法を用いることも可能である．

2.3　DBL アルゴリズムの収束性

DBL アルゴリズムは，各ステップにおいて，$F(\theta_t,\theta_t)$ を小さくするという保証はない．しかし，下記の条件のもとでは収束性を保証できる．

定理 1.

θ_t が DBL アルゴリズムに従って更新されたとする．このとき，以下は θ_t が収束するための十分条件となる．

$$
\begin{aligned}
&F(\theta_{t+1},\theta_t) < F(\theta_{t-1},\theta_t) \text{ であるとき，}\\
&F(\theta_t,\theta_{t+1}) < F(\theta_t,\theta_{t-1}) \text{ が成り立つ．}
\end{aligned}
\tag{4}
$$

定理 1 は，$F(\theta_{t+2},\theta_{t+1}) < F(\theta_t,\theta_{t-1})$ を導くことで証明できるが，詳細は省略する．条件 (4) は，KL 擬距離の差分 $KL[q_t|p_{t+1}] - KL[q_t|p_t]$ が正であるとき，KL 擬距離において分布を入れ替えた $KL[p_{t+1}|q_t] - KL[p_t|q_t]$ も正となれば満たされる．ここで，$q_t \equiv p(\mathbf{v}'|\mathbf{v},\theta_t)r(\mathbf{v})$, $p_t \equiv p(\mathbf{v}|\mathbf{v}',\theta_t)r(\mathbf{v}')$ である．例えば，分布 p_t, p_{t+1}, q_t が異なる平均，等しい共分散行列を持つガウス分布のとき，条件 (4) は常に成り立つ．

3　Contrastive Divergence Learning (CDL) との関係

DBL は，restriceted Boltzmann machine (RBM) の学習において CDL と密接な関係を持つことを以下で示す．

RBM は，観測変数 $\mathbf{v} \in \{0,1\}^n$ と隠れ変数 $\mathbf{h} \in \{0,1\}^m$ の分布として以下のように定義される．

$$
p(\mathbf{v},\mathbf{h}|\theta) = \frac{1}{Z}\exp\left(-E(\mathbf{v},\mathbf{h};\theta)\right), \tag{5}
$$

ここで Z は正規化定数である．T をベクトルの転置を表すものとすると，$E(\mathbf{v},\mathbf{h};\theta) \equiv -\mathbf{v}^{\mathrm{T}}W\mathbf{h} - \mathbf{b}^{\mathrm{T}}\mathbf{v} - \mathbf{c}^{\mathrm{T}}\mathbf{h}$ と定義され，$W, \mathbf{b}, \mathbf{c}$ はそれぞれ $n\times m$ 行列，n 次元ベクトル，m 次元ベクトルである．学習すべきパラメータは $\theta = \{W,\mathbf{b},\mathbf{c}\}$ である．同時分布 $p(\mathbf{v},\mathbf{h}|\theta)$ は評価することが難しいが，条件付き分布 $p(\mathbf{v}|\mathbf{h},\theta), p(\mathbf{h}|\mathbf{v},\theta)$ はそれぞれ要素ごとに独立な二項分布の積となる．これらの条件付き分布に基づき，ギブスサンプリングを行うことによって同時分布からのサンプリングが可能となる．ギブスサンプラーを $p_G(\mathbf{v}|\mathbf{v}',\theta) = \sum_{\mathbf{h}'} p(\mathbf{v}|\mathbf{h}',\theta)p(\mathbf{h}'|\mathbf{v}',\theta)$ とし，初期分布 $Q_0(\mathbf{v}|\theta)$ を $r(\mathbf{v})$ としたときの t ステップ後の分布を $Q_t(\mathbf{v}|\theta) \equiv \sum_{\mathbf{v}'} p_G(\mathbf{v}|\mathbf{v}',\theta)Q_{t-1}(\mathbf{v}'|\theta)$ と書く．CDL では次のコスト関数の最小化を考える．$CD(\theta) \equiv KL[Q_0(\mathbf{v})|Q_\infty(\mathbf{v}|\theta)] - KL[Q_1(\mathbf{v}|\theta)|Q_\infty(\mathbf{v}|\theta)]$．ここで，$Q_\infty(\mathbf{v}|\theta) \equiv \lim_{t\to\infty} Q_t(\mathbf{v}|\theta) = \sum_{\mathbf{h}} p(\mathbf{v},\mathbf{h}|\theta)$ である．このコスト関数自体は評価することが困難であるが，パラメータ微分から確率勾配による更新則を近似的に求めることができる．確率勾配法による更新則は，ある正のステップサイズを η_t，$\Delta\theta = -\left.\frac{\partial CD(\theta)}{\partial\theta}\right|_{\theta=\theta_t}$ として，$\theta_{t+1} = \theta_t + \eta_t\Delta\theta$ の形で与えられる．特に学習するパラメータとして W の (i,j) 要素 w_{ij} を選んだとき，

$$
\begin{aligned}
-\frac{\partial CD(\theta)}{\partial w_{ij}} &= \langle v_i h_j\rangle_{r(\mathbf{v})p(\mathbf{h}|\mathbf{v},\theta_t)} - \langle v_i h_j\rangle_{Q_1(\mathbf{v}|\theta_t)p(\mathbf{h}|\mathbf{v},\theta_t)}\\
&\quad - \frac{\partial H(Q_1(\mathbf{v}|\theta_t))}{\partial w_{ij}}
\end{aligned}
\tag{6}
$$

となる． ここで，$H(p(\mathbf{v})) \equiv -\sum_{\mathbf{v}} p(\mathbf{v})\log p(\mathbf{v})$ は分布 $p(\mathbf{v})$ のエントロピーを表す．CDL では計算困難な上記の第三項を無視し，以下の更新則を用いる．

$$
\Delta w_{ij} = \langle v_i h_j\rangle_{r(\mathbf{v})p(\mathbf{h}|\mathbf{v},\theta_t)} - \langle v_i h_j\rangle_{Q_1(\mathbf{v}|\theta_t)p(\mathbf{h}|\mathbf{v},\theta_t)} \tag{7}
$$

一方，DBL アルゴリズムにおけるステップ 2 の更新則も確率勾配法によって得られる．コスト関数 (3) の微分から更新則を導くと，これは近似なしに式 (7) と一致する．この更新則の一致は DBL と CDL のコスト関数の間に成り立つ次の関係式からも理解できる．

定理 2.

マルコフ連鎖 $p(\mathbf{v}|\mathbf{v}',\theta)$ の定常分布 $p_\infty(\mathbf{v}|\theta)$ が詳細釣り合い条件を満たすとき，以下が成り立つ．

$$
F(\theta,\theta) = CD(\theta) + KL[Q_1(\mathbf{v}|\theta)|r(\mathbf{v})] \tag{8}
$$

証明は省略する．KL 擬距離や CD が非負の値をとる [1] ので，定理 2 は $F(\theta,\theta)$ が $CD(\theta)$ や $KL[p_1(\mathbf{v}|\theta)|r(\mathbf{v})]$ の上限を与えることを示している．最後に RBM のような隠れ状態を含むマルコフモデルにおいて真の分布 $r(\mathbf{v})$ に対して (狭義の) 詳細釣り合い条件が成り立つ必要十分条件を示す．

定理 3.

$p(\mathbf{v},\mathbf{h}|\theta)$ の条件付き分布 $p(\mathbf{v}|\mathbf{h},\theta)$ と $p(\mathbf{h}|\mathbf{v},\theta)$ から構成されるマルコフ連鎖 $p_G(\mathbf{v}|\mathbf{v}',\theta) = \sum_{\mathbf{h}} p(\mathbf{v}|\mathbf{h},\theta)p(\mathbf{h}|\mathbf{v}',\theta)$ が唯一の定常分布をもつとする．このとき，以下が成り立つ．

$$
\begin{aligned}
&r(\mathbf{v}) = \sum_{\mathbf{h}} p(\mathbf{v},\mathbf{h}|\theta)\\
\Leftrightarrow\quad &p(\mathbf{v}'|\mathbf{h},\theta)p(\mathbf{h}|\mathbf{v},\theta)r(\mathbf{v}) = p(\mathbf{v}|\mathbf{h},\theta)p(\mathbf{h}|\mathbf{v}',\theta)r(\mathbf{v}')
\end{aligned}
$$

4　まとめ

本研究では，マルコフ連鎖の定常分布を学習する方法として DBL を提案し，DBL が CDL と密接な関係をもつことを示した． 今後，DBL の学習速度，収束性の改善を検討する予定である．

参考文献

[1] G. E. Hinton (2002) “Training Products of Experts by Minimizing Contrastive Divergence.” Neural Computation, **14(8)**, 1771–1800.