クラスターBPによるKikuchi近似の計算

佐藤泰介
東京工業大学情報理工学研究科
email: sato@mi.cs.titech.ac.jp

1 はじめに

ベイジアンネット（*Bayesian net*） [2] の周辺分布の計算法には 単結合（*singly connected*）グラフに使える Pearl のアルゴリズム [7] と複結合（*multiply connected*）グラフでも使える結合木（*junction tree*） アルゴリズム [4] がある。ともに頂点間のメッセージ交換に基づくアルゴリズムであり、まとめて **BP** (*belief propagation*) と呼ばれている。特に BP をループのあるベイジアンネットまで適用した場合 **loopy BP** と呼ばれている。

Loopy BP は誤り訂正符号の分野で驚異的性能を誇る Turbo 復号アルゴリズムや LDPC 復号アルゴリズムと一致する事が近年分かり [5]、 多変量分布の周辺分布の効率的近似計算法として注目を集めている。 現在 loopy BP は統計力学に於ける変分自由エネルギーの近似法の一つである Bethe 近似の停留解を与える計算法として理解されている。更に良い近似としてはより一般的な Kikuchi 近似 [8] があり、Kikuchi 近似の停留解を与える計算法として GBP[8] や CCCP[9] がすでに提案されている。 しかしながら GBP も CCCP もアルゴリズムとして複雑であり、どのような場合でも効率的であるとは言えない。

Aji ら [1] は GBP より近似精度は落ちるが計算効率の良い方法として、loopy BP を変数クラスターを頂点とする結合グラフ (*junction graph*) に適用したが、それが Kikuchi 近似の停留解を計算している保証はなかった。本論文では [1] の問題点を克服するため、**クラスターグラフ**を導入し、クラスターグラフ上の BP、即ち**クラスター BP** を提案する。そしてクラスターグラフが”**tree 条件**”を満たす時クラスター BP は Kikchi 近似の停留解を計算する。紙面の都合上、詳しい説明及びクラスターグラフ上の CCCP アルゴリズムの導出は省いた[1]。

[1] http://sato-www.cs.titech.ac.jp/sato/Cluster-BP.ps に本論文のより詳しい説明がある。そこでは本論文に於ける「クラスターグラフ」を「一般化クラスターグラフ」、「クラスター BP」を「一般化クラスター BP」と呼んでいる。 また Bethe 近似、Kikuchi 近似の説明は [8] に依っている。

2 準備

我々は計算の対象である離散的な n 次元確率変数ベクトル $X = (X_1, \ldots, X_n)$ を「系」と呼び、その同時確率分布 $p(x) = p(x_1, \ldots, x_n)$ が以下の条件を満たす事を仮定する。

- $p(x)$ は常に正、即ち $p(x) > 0$ for any x
- $p(x)$ は以下のように**ポテンシャル関数** $\psi_\alpha(x_\alpha)$ の積として表される

$$p(x) = \kappa \prod_{\alpha \in \mathbf{P}} \psi_\alpha(x_\alpha) \qquad (1)$$

where

$$\kappa^{-1} = \sum_x \prod_{\alpha \in \mathbf{P}} \psi_\alpha(x_\alpha)$$

ここで κ, κ' などは正規化定数を表す。 α はベクトル $(1, \ldots, n)$ の部分ベクトルであり、 $\mathbf{P}$ はそのような部分ベクトルの集合である。 x_α は α に対応する $(x_1, \ldots, x_n)$ の部分ベクトルを表す。 例えば $x = (x_1, x_2, x_3, x_4)$ に対し $\alpha = (2,4)$ ならば $x_\alpha = (x_2, x_4)$ となる。便宜上部分ベクトル α を集合と同一視する。従って $\alpha = (2,4)$ は $\alpha = \{2,4\}$ でもある。α の成分は順序付けられているのでこのようにしても混乱は起きない。 以後 $\{1, \ldots, n\}$ の部分集合を **クラスター**、$\mathbf{P}$ の元を**ポテンシャルクラスター**と呼ぶ。α, β がクラスターの時 $\alpha \cap \beta$、 $\alpha \cup \beta$、 $\alpha \setminus \beta$ もクラスターである。$\alpha = \{1,2,3\}$、 $\beta = \{2,3,4\}$ ならば $x_{\alpha\cap\beta} = x_{\{2,3\}} = (x_2, x_3)$、$x_{\alpha\setminus\beta} = x_{\{1\}} = x_1$ などとなる。 また記法上の注意として $A \subset B$ は A が B

の真部分集合である事を意味し、$A \subseteq B$ は $A \subset B$ または $A = B$ を意味する。

さてここで統計力学に習い、試験分布 $b(\cdot)$ を変分関数とする変分自由エネルギー

$$\begin{aligned} F(b) &\stackrel{\text{def}}{=} (\text{系の平均エネルギー}) \quad (2) \\ &\quad -(\text{系のエントロピー}) \\ &= \sum_x b(x)E(x) - S \quad (3) \end{aligned}$$

を考えると[2] $F(b)$ は $b(x) = p(x)$ の時最小値を取る。我々の目標は、式 (1) で定義される分布 $p(x)$ が与えられた時、式 (3) の変分自由エネルギー $F(b)$ に対し $F(b) \approx F_K(\{b_\alpha\})$ であるような *Kikuchi* **近似** $F_K(\{b_\alpha\})$ を定義し、$p(x) = \text{argmin}_b F(b)$ から導かれる $p(x)$ の周辺分布 $\{p_\alpha(x_\alpha)\}$ に関する近似式

$$\{p_\alpha\} \approx \text{argmin}_{\{b_\alpha\}} F_K(\{b_\alpha\})$$

を利用して、$\text{argmin}_{\{b_\alpha\}} F_K(\{b_\alpha\})$ を計算する事により $\{p_\alpha\}$ を近似的に求める事である。但し $b_\alpha(x_\alpha)$ は $b(x)$ の α の周辺分布である。

3 クラスターと重複数

ポテンシャルクラスター $\mathbf{P}$ に対し以下の性質を持つ集合 $\mathbf{U}$ を導入する。

- $\mathbf{P}$ の任意の元 α に対し、それを覆う $\mathbf{U}$ の（集合の包含関係の意味で）最小元 $\beta \in \mathbf{U}$ がある。即ち $\alpha \subseteq \beta$ 且つ、任意の $\alpha' \in \mathbf{U}$ に対し $\alpha \subseteq \alpha'$ ならば $\beta \subseteq \alpha'$

$\mathbf{U}$ は集合の包含関係に関して半順序集合をなしている。$\mathbf{U}$ の元を**一般化** *Kikuchi* **クラスター**と呼ぶ。元々の Kikuchi クラスター集合は交わりに関して閉じている（$\alpha, \beta \in \mathbf{U}$ ならば $\alpha\beta \in \mathbf{U}$）事が要求されているが、ここではそのような事は求められていない。本論文で以後簡単の為 $\mathbf{U}$ の極大元集合と $\mathbf{P}$ が一致する場合のみを考える。次に $\mathbf{U}$ の各元 α に対し**重複数** a_α を以下により決める。

$$\begin{cases} a_\alpha \stackrel{\text{def}}{=} 1 \ \text{if } \alpha \text{ が極大} \\ a_\beta \stackrel{\text{def}}{=} 1 - \sum_{\alpha:\alpha\supset\beta, \alpha\in\mathbf{U}} a_\alpha \quad \text{その他} \end{cases}$$

[2] 系のエネルギー $E(x)$ は $E(x) = -\sum_{\alpha\in\mathbf{P}} \ln\psi(x_\alpha)$、エントロピーは $S = -\sum_x b(x)\ln b(x)$ で与えられる。

この定義から

$$\sum_{\alpha:\alpha\in\mathbf{U}, \alpha\supseteq\beta} a_\alpha = 1 \ \text{ for } \forall\beta \in \mathbf{U}$$

が成立する。この性質を利用して系のエントロピーの近似を行なう。まずクラスター $\alpha \in \mathbf{U}$ のエントロピー S_α を

$$S_\alpha \stackrel{\text{def}}{=} -\sum_{x_\alpha} b_\alpha(x_\alpha)\ln b_\alpha(x_\alpha)$$

と置く。次に S^+_α を α の小さい方から帰納的に定める。

$$\begin{cases} S^+_\alpha \stackrel{\text{def}}{=} S_\alpha \quad \text{if } \alpha \text{ が } \mathbf{U} \text{ の極小元} \\ S^+_\alpha \stackrel{\text{def}}{=} S_\alpha - \sum_{\beta:\beta\subset\alpha, \beta\in\mathbf{U}} S^+_\beta \quad \text{o.w.} \end{cases}$$

すると

$$\begin{aligned} \sum_x b(x)E(x) &= \sum_{\alpha\in\mathbf{U}} a_\alpha \sum_{x_\alpha} b_\alpha(x_\alpha)E_\alpha(x_\alpha) \\ S &\approx \sum_{\beta\in\mathbf{U}} S^+_\beta \\ &= \sum_{\alpha\in\mathbf{U}} a_\alpha S_\alpha \end{aligned}$$

が導ける。但し $E_\alpha(x_\alpha)$ は $E_\alpha(x_\alpha) = \sum_{\beta:\beta\in\mathbf{P}, \beta\subseteq\alpha}(-\ln\psi_\beta(x_\beta))$ により定められるクラスター α のエネルギーである。

これら二つの式を式 (3) に代入すると $F(b)$ の近似式が得られ、さらにそれを同値変形する事により変分自由エネルギー $F(b)$ の $\mathbf{U}$ **に基づく一般化** *Kikuchi* **近似** $F'_K(b)$

$$F'_K(\{b_\alpha\}) \stackrel{\text{def}}{=} \sum_{\alpha\in\mathbf{U}} a_\alpha \sum_{x_\alpha} b_\alpha(x_\alpha)\ln\left(\frac{b_\alpha(x_\alpha)}{\phi_\alpha(x_\alpha)}\right) \quad (4)$$

where

$$\phi_\alpha(x_\alpha) = \prod_{\beta:\beta\in\mathbf{P}, \beta\subseteq\alpha} \psi_\beta(x_\beta) \quad (5)$$

を得る。$\mathbf{U}$ が交わりに関して閉じている時の $F'_K(\{b_\alpha\})$ が本来の Kikuchi 近似である。$F_K(\{b_\alpha\})$ の停留解（最小値）を与える $b_\alpha(x_\alpha)$ は $p(x)$ の正確な周辺分布 $p_\alpha(x_\alpha)$ の良い近似を与える事が期待出来る。

4 クラスターグラフ

$F'_K(\{b_\alpha\})$ の停留解を BP のようなメッセージ交換で計算するため、**クラスターグラフ**を導入する。クラスターグラフは結合木や Aji らの結合グラフ [1] を特殊な場合として含む。

一般化 Kikuchi クラスター集合 $\mathbf{U}$ に基づくクラスターグラフ $\mathcal{G}_U$ とは頂点と辺が $\mathbf{U}$ の元によりラベル付けされた無向グラフで以下の条件を満たすものを言う。

- $\mathbf{U}$ の極大元は丁度 1 回頂点のラベルとして $\mathcal{G}_U$ に出現している。
- 頂点 v_1, v_2 がそれぞれクラスター α_1、α_2 によりラベル付けされている時、v_1 と v_2 を結ぶ辺（もしあれば）は、$\beta \subseteq \alpha_1 \cap \alpha_2$、$\beta \in \mathbf{U}$ でラベル付けされている。

クラスターグラフ上の BP のようなメッセージ交換が Kikuchi 近似の停留解を計算する事を保証する為 **tree 条件**を導入する。

> $\mathcal{G}_U$ が *tree* **条件** を満たす
> iff
> 任意の $\alpha \in \mathbf{U}$ について、α を含むラベルでラベル付けされた頂点、辺よりなる $\mathcal{G}_U$ の部分グラフが tree を成す

図 1 の左のクラスターグラフは $\mathbf{U} = \{124, 234, 134, 24, 34, 14, 4\}$ のクラスターグラフである。4 を含む頂点と辺がループを成すので tree 条件を満たしていないが、4 を辺から削除した右のグラフは tree 条件を満す。

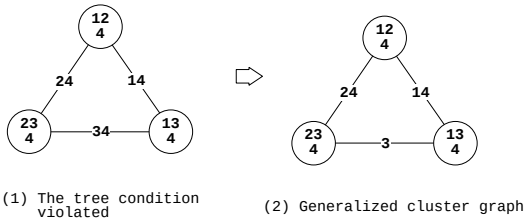

図 1: tree 条件を満たすクラスターグラフへの変換

5 クラスター BP の導出

$\alpha \in \mathbf{U}$ によりラベル付けされているクラスターグラフ $\mathcal{G}_U$ の頂点の数を $V(\alpha)$、辺の数を $E(\alpha)$ で表す。$\mathbf{U}$ の半順序関係を使った帰納法により我々は以下の事を証明出来る。

Proposition 5.1 $\mathcal{G}_U$ が *tree* 条件を満たすなら、任意の重複数 $a_\alpha \in \mathbf{U}$ について

$$a_\alpha = V(\alpha) - E(\alpha)$$

これより、一般化 Kikuchi 近似は以下のように変形される。

$$\begin{aligned} F'_K(\{b_\alpha\}) &= \sum_{v \in V} \sum_{x_{\underline{v}}} b_v(x_{\underline{v}}) \ln \left(\frac{b_v(x_{\underline{v}})}{\phi_{\underline{v}}(x_{\underline{v}})} \right) \\ &\quad - \sum_{e \in E} \sum_{x_{\underline{e}}} b_e(x_{\underline{e}}) \ln b_e(x_{\underline{e}}) \end{aligned}$$

但し V、E はそれぞれ $\mathcal{G}_U$ の頂点の全体、辺の全体であり、頂点 $v \in V$ をラベル付けしているクラスターを $\underline{v}$ で、辺 $e \in E$ をラベル付けしているクラスターを $\underline{e}$ で表している。我々は F'_K を極小化する $\{b_v(x_{\underline{v}}), b_e(x_{\underline{e}})\}$ を、確率制約 $\sum_{x_{\underline{v}}} b_v(x_{\underline{v}}) = 1$ と周辺分布間の一貫性制約 $\sum_{x_{\underline{v_1} \setminus \underline{e}}} b_{v_1}(x_{\underline{v_1}}) - b_e(x_{\underline{e}})$ の下に求める。その為

$$\begin{aligned} \mathcal{C} &= \sum_{e \in E} \sum_{x_{\underline{e}}} \left\{ \lambda^e_{v_1}(x_{\underline{e}}) \left(\sum_{x_{\underline{v}^1 \setminus \underline{e}}} b_{v_1}(x_{\underline{v_1}}) - b_e(x_{\underline{e}}) \right) \right. \\ &\quad \left. + \lambda^e_{v_2}(x_{\underline{e}}) \left(\sum_{x_{\underline{v}^2 \setminus \underline{e}}} b_{v_2}(x_{\underline{v_2}}) - b_e(x_{\underline{e}}) \right) \right\} \\ &\quad + \sum_{v \in V} \lambda_v \left(\sum_{x_{\underline{v}}} b_v(x_{\underline{v}}) - 1 \right) \end{aligned} \tag{6}$$

なる制約項 $\mathcal{C}$ を F'_K に加えた Lagrangian$\mathcal{L} = F'_K + \mathcal{C}$ の停留点を求める。ここで $\{\lambda^e_{v_1}(x_{\underline{e}}), \lambda^e_{v_2}(x_{\underline{e}}), \lambda_v\}$ は Lagarange 乗数である。実際に $\mathcal{L}$ を微分して計算を行い、

$$m_{ij}(x_{\underline{i} \cap \underline{j}}) = \kappa \; e^{-\lambda^{ij}_{i}(x_{\underline{i} \cap \underline{j}})}$$

$$m_{ij}(x_{\underline{ij}}) = \kappa \sum_{x_{\underline{i}\backslash\underline{ij}}} \phi_i(x_{\underline{i}}) \prod_{k\in N(i)\backslash j} m_{ki}(x_{\underline{ik}}) \quad (7)$$

$$b_i(x_{\underline{i}}) = \kappa' \, \phi_i(x_{\underline{i}}) \prod_{k\in N(i)} m_{ki}(x_{\underline{ik}}) \quad (8)$$

図 2: **クラスター BP**

と置くと図 2 の**クラスター BP** に到達する。$m_{ij}(x_{\underline{ij}})$ は頂点 i から頂点 j に向かうメッセージであり、 $x_{\underline{ij}}$ の $\underline{ij}$ は i と j を結ぶ辺をラベル付けしている $\mathbf{U}$ のクラスターである。

ここで導入したクラスター BP は GBP より狭いが、Pearl の BP や接合木の BP、Aji らの結合グラフ上の BP[1]、ファクターグラフ上の sum-product アルゴリズム [3] を包含する、(一般化された) Kikuchi 近似を計算する反復アルゴリズムである。

Pakzad らは [6] でクラスター集合 $\mathbf{R}$ を導入し、 $\mathbf{R}$ の集合の成す半順序関係を表す Hassse 図上のメッセージ交換を論じている。変分自由エネルギーの近似程度の目安としてクラスターサイズ n に注目し、 サイズ n 以下の任意の集合 $C_k = \{i_1, \ldots, i_k\} (k \leq n)$ について、条件 $\mathbf{B}n$: $\sum_{\alpha\in\mathbf{R}:\alpha\supseteq C_k} a_\alpha = 1$ を付ける事により近似程度の制御を計っている [6]。 一般化クラスター BP との比較で言えば、条件 $\mathbf{B}n$ は一般化クラスター BP で成り立っている条件、 即ち任意の $\beta \in \mathbf{U}'$ に対し $\sum_{\alpha:\alpha\in\mathbf{U}',\alpha\supseteq\beta} a_\alpha = 1$、とは独立である。その為一般化クラスターグラフでも Pakzad らのアプローチで認められないものがある。例えばクラスター集合 $\{123, 234, 2\}$ の Hasse 図は tree 条件を満たす一般化クラスターグラフである。 しかし $a_{123} = a_{234} = 1$ なので $\sum_{\alpha\supseteq 23} a_\alpha = 1 + 1 = 2$ となり、$n = 2$ の時の条件 $\mathbf{B}2$ が満たされないので Pakzad らのアプローチでは認められない。また彼らのメッセージが「上から下へ」の一方向である点もクラスター BP と異なる。

6 Acknowledgment

本論文は統計力学、AI、符号理論の境界線上にあり、様々な人との討論を通じて生まれたものである。荒木純道氏には統計力学や通信についての知識を授けて頂き、渡辺治氏には計算論的観点からの視点を教えて頂いた。特に渋谷智治氏には LDPC など誤り訂正符号について具体的に教えて頂いた。記して感謝を記したい。

References

[1] A.M. Aji and R.J. McEliece. The generalized distributive law. *IEEE Trans. Inform. Theory*, 46(2):325–343, 2000.

[2] E. Castillo, J. M. Gutierrez, and A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer-Verlag, 1997.

[3] F.R. Kschischang, B.J. Frey, and H.A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transaction on Information Theory*, 47(2):498–519, 2001.

[4] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their applications to expert systems. *Journal of the Royal Statistical Society, B*, 50:157–224, 1988.

[5] D.J.C. MacKay, R.J. McEliece, and J.-F. Cheng. Turbo decoding as an instance of pearl's "belief propagation" algorithm. *IEEE Journal of Selected Areas of Communication*, pages 140–152, 1998.

[6] P. Pakzad and V. Anantharam. Belief Propagation and Statistical Physics. In *Proceedings of 2002 Conference on Information Sciences and Systems*, 2002.

[7] J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29:357–369, 1986.

[8] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Bethe free energy, kikuchi approximations, and belief propagation algorithms. Technical Report MERL TR2001-16, Mitsubishi Electric Research Laboratories, 2001.

[9] A.L. Yuille. CCCP algorithms to minimize the bethe and kikuchi free energies: Convergent alternatives to belief propagation. In *NIPS2001*, pages –, 2001.