Statistical-Mechanical Approach to Probabilistic Information Processing

Nov. 22, 2005

計算限界

確率伝播法とその可能性について[1]

渡辺 治（東京工業大学）

本日の話題

1. 確率伝播法 (Belief Propagation Method)
2. メッセージ伝播アルゴリズムの特徴と可能性について
3. アルゴリズムの平均的解析のモデル (Planted Solution Model)

例題として：グラフの二分割問題 Graph Bisection Problem

[1] このファイルは発表時に使用したプレゼンテーション用ファイルをＡ４版に変更したものです．スタイル等が悪くて見にくい部分がありますがご勘弁下さい．

1．確率伝播法 (Belief Propagation Method) について

一言でいうと

確率変数 $X_1, ..., X_n$（各々の定義域を $A_1, ...A_n$ とする）により決まる系 G において，次のような周辺確率を求める方法である．

$$p(a) \;=\; \Pr_G[\, X_1 = \boxed{a} \,] \;=\; \sum_{a_2 \in A_2,\ ...,\ a_n \in A_n} \Pr[\, X_1 = \boxed{a} \wedge X_2 = a_2 \wedge \cdots X_n = a_n \,].$$

ただし

確率変数間の依存関係が単純で，同時確率分布を分解して表わすことができる場合に限る $\Longrightarrow$ ベイジアン・ネットワーク

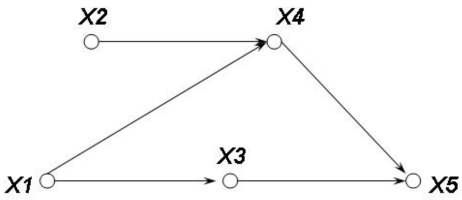
歴史

- AI の分野で Pearl が 1982 年に提唱したもの．
- 統計力学の研究者 MacKay らが LDPC 符号の復号法に使用して注目を集めた．それ以後，統計力学＆情報理論を中心に研究が進んでいる．
- 最近では survey propagation なるものも登場している．

上記の周辺確率よりも，もっと一般的には次のような確率を求める（求めたい）ことが多い．

$$\begin{aligned} &\Pr_G[\, X_1 = a \,|\, X_n = b] \\ &= \frac{\displaystyle\sum_{a_2 \in A_2,\ ...,\ a_{n-1} \in A_{n-1}} \Pr[\, X_1 = \boxed{a} \wedge X_2 = a_2 \wedge \cdots X_n = b \,]}{\displaystyle\sum_{a_1 \in A_1,\ a_2 \in A_2,\ ...,\ a_{n-1} \in A_{n-1}} \Pr[\, X_1 = a_1 \wedge X_2 = a_2 \wedge \cdots X_n = b \,]} \end{aligned}$$

ベイジアン・ネットワーク：
同時確率分布の分解法を示した確率変数間の依存グラフ



$$p(a_1, a_2, a_3, a_4, a_5) \;=\; p(a_1)p(a_2)p(a_3|a_1)p(a_4|a_1, a_2)p(a_5|a_3, a_4)$$

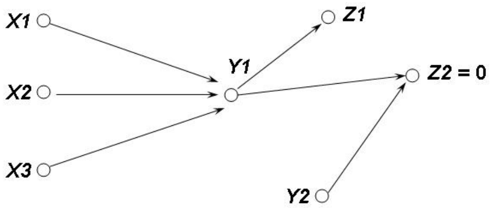
ただし，たとえば次のように略記する：

$$\mathrm{Pr}_G[\, X_2 = a_2 \,\&\, X_4 = a_4 \,] \;=\; p(a_2, a_4)$$

注意！（矢印がない関係はすべて独立）

$$\begin{aligned} p(a_1, a_2, a_3, a_4, a_5) &= \underline{p(a_1)}p(a_2)\underline{p(a_3|a_1)}p(a_4|a_1, a_2)p(a_5|a_3, a_4) \\ &= p(a_2)p(a_1, a_3)p(a_4|a_1, a_2)p(a_5|a_3, a_4) \\ &= p(a_1, a_2, a_3, a_4)p(a_5|a_3, a_4) \\ &= p(a_1, a_2, a_3, a_4, a_5) \end{aligned}$$

BP の計算方法：
確率変数間のメッセージ伝播により各確率変数の belief を計算する．



毎ステップ，各頂点において，次のような計算を行う：

以下，頂点 Y_1 と，Y_1 の取りうる値 a に注目．
（準備）

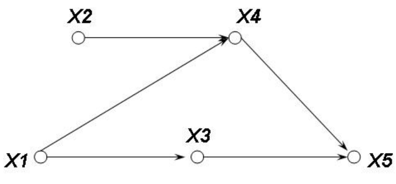
- $\pi_{Y_1}(a) = \Sigma_{\boldsymbol{b}}\ \Pr[Y_1 = a | \boldsymbol{X} = \boldsymbol{b}] \cdot (\,\pi_{X_1\to Y_1}(b_1)\cdot\pi_{X_2\to Y_1}(b_2)\cdot\pi_{X_3\to Y_1}(b_3)\,)$.
- $\lambda_{Y_1}(a) = \lambda_{Z_1\to Y_1}(a)\cdot\lambda_{Z_2\to Y_1}(a)$.

（メッセージの計算）
親へ（例：$Y_1 \longrightarrow Z_1$）：　$\pi_{Y_1,Z_1}(a) = \pi_{Y_1}(a)\cdot\lambda_{Z_2\to Y_1}(a)$.
子へ（例：$Y_1 \longrightarrow X_1$）：　$\lambda_{Y_1\to X_1}(b) \approx \Pr[\,X_1 = b\,]$ の一部
（Y_1 への X_1 以外の情報により得られる値）

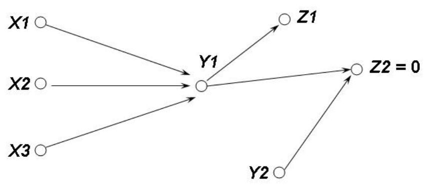
（現時点での $Y_1 = a$ に対する belief の計算）
$\mathrm{Bel}_{Y_1}(a) = \alpha\cdot\lambda_{Y_1}(a)\pi_{Y_1}(a)$.

これまでにわかっていること：

定理．
1. 依存グラフに閉路がなければ正しく周辺確率を計算する．
2. その場合の計算時間は $\mathcal{O}(n)$（n は頂点数）である．



閉路あり



閉路なし

でも，経験的に...　閉路があっても大抵うまくいく場合がある！
1. LDPC 符号（Gallager 符号）の復号化
2. Turbo 符号の復号化
3. 画像修復，様々な制約充足問題

２．ＢＰ（メッセージ伝播法）の特徴と可能性

私の直感：

$$\text{確率伝播法} \iff \text{乱択局所探索アルゴリズム}$$

補足

- 乱択局所探索アルゴリズムにもいろいろあり，その性質も大きく異なる．上記の対応は「貪欲型」に対するもの.（もっと正確には，決定にランダム性を入れて多少ソフト化したもの.）

- 単純に乱択化しても効率が悪くなってしまう場合がある．

  比ゆ： 5 回サイコロを投げた和が 15 の元で，すべて 3 が出る確率 P_0

$$\underset{\text{計算で求まる}}{P_0 \;=\; \cdots} \iff \underset{\text{多量のサンプルが必要}}{P_0 \;=\; \frac{\text{すべて 3 の回数}}{\text{和が 15 の回数}}}$$

- 「貪欲型」が有効な場合に使える？(収束性がよい場合)

- もちろん（？）最悪時はダメ，平均的によい $\Longrightarrow$ 平均時解析

３．平均時解析の例

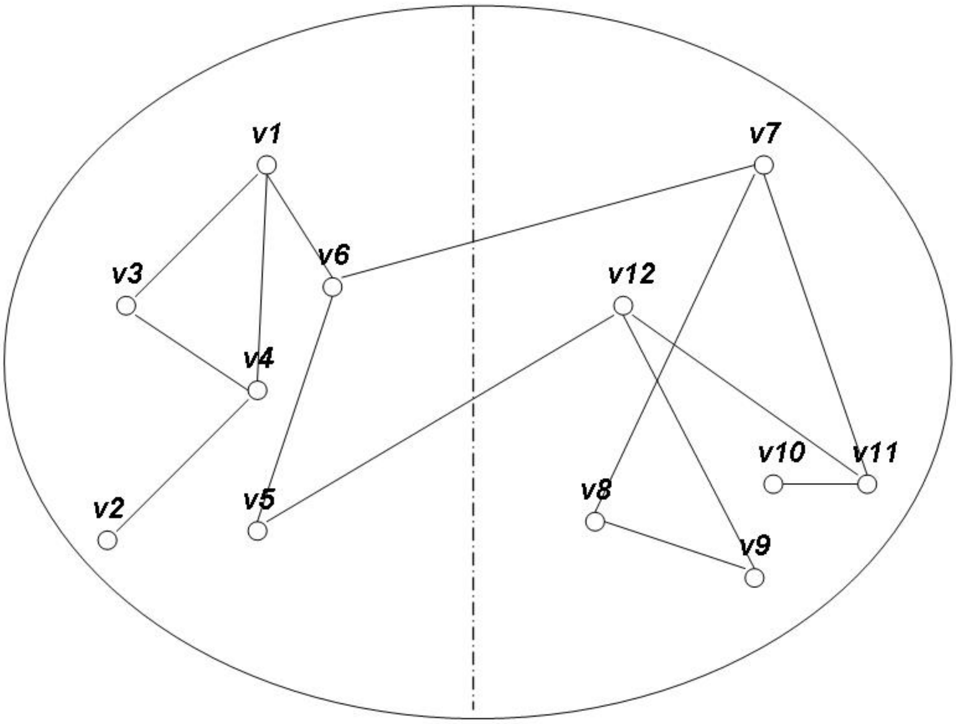
例題： グラフの二分割問題 Graph Bisection Problem

Graph Bisection Problem

Input: An undirected graph $G = (V, E)$, where $|V| = 2n$ for some n.

Task: Find an *equal size* partition V_0 and V_1 of V with min. # of cut edges.

Remark. Find one of them if there are several solutions.



知られていること：

- この問題はＮＰ-困難.
- 平均で多項式時間のアルゴリズムが知られている.
  その入力グラフのモデルとして planted solution model が導入された.

  > (Boppana, 1987): deterministic polynomial-time but not so simple
  > (Jerrum and Sorkin, 1998): randomized $\mathcal{O}(n^2)$ time
  > (Condon and Karp, 2001): randomized $\mathcal{O}(n+m)$ time
  > (F. M^cSherry, 1999, etc): simplified spectral method

ＢＰによるアルゴリズムの導出

確率変数

C_i $(1 \le i \le 2n)$: 頂点 v_i が V_0, V_1 のどちらに属するか？

$E_{i,j}$ $(1 \le i < j \le 2n)$: $E_{i,j} = 1 \Leftrightarrow v_i, v_j$ 間に辺が存在する．

計算の目標

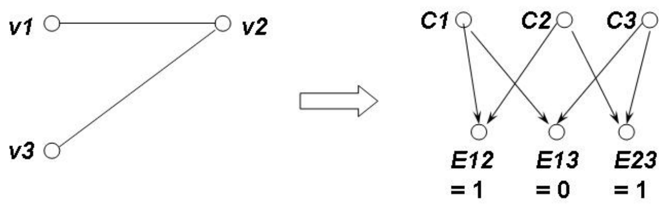
与えられたグラフ G に対し，ＢＰを用いて次の周辺確率を求める．

$$p_i \;=\; \Pr[\, C_i = 1 \,|\, C_1 = 1 \;\wedge\; \bigwedge_{i,j} E_{i,j} = e_{i,j} \,]$$

$$\text{ただし，}\quad e_{i,j} \;=\; \begin{cases} 1, & G \text{ は辺 } (v_i, v_j) \text{ を持つ，} \\ 0, & G \text{ は辺 } (v_i, v_j) \text{ を持たない．} \end{cases}$$

その $p_i > 0.5$ か否かで C_i を決める．

ベイジアン・ネットワーク

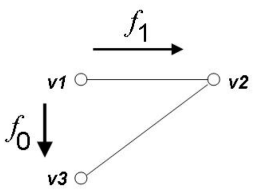


得られたアルゴリズム

非常に単純なメッセージ伝播アルゴリズム

program `PseudoBP`$(G = (V, E), p, r)$;
  set all x_i to 0;
  repeat MAXSTEP times **do** {
    $x_1 \leftarrow +\infty$;
    for each $i \in V$,
      $$x_i \ \leftarrow \ \sum_{j \in N(i)} f_1(x_j) + \sum_{j \notin N(i)} f_0(x_j);$$
    if all beliefs are stabilized **then break**;
  }
  output$(x_1, ..., x_{2n})$; // Classification is $(\mathrm{sign}(x_1), ..., \mathrm{sign}(x_{2n}))$.
end-program

補足：



1. 関数 f_0, f_1 の定義：

$$f_1(z) = \begin{cases} h_1 \cdot \theta_1, & \text{if } \theta_1 < z, \\ h_1 \cdot z, & \text{if } -\theta_1 \le z \le \theta_1, \\ h_1 \cdot (-\theta_1), & \text{if } z < -\theta_1; \end{cases} \quad f_0(z) = \begin{cases} -h_0 \cdot \theta_0, & \text{if } \theta_0 < z, \\ -h_0 \cdot z, & \text{if } -\theta_0 \le z \le \theta_0, \\ -h_0 \cdot (-\theta_0), & \text{if } z < -\theta_0, \end{cases}$$

ただし,

$$a_0 \ = \ \frac{1-p}{1-r}, \quad a_1 \ = \ \frac{p}{r}, \quad h_0 \ = \ \left|\frac{a_0 - 1}{a_0 + 1}\right|, \quad h_1 \ = \ \left|\frac{a_1 - 1}{a_1 + 1}\right|, \quad \theta_0 \ = \ \left|\frac{\ln a_0}{h_0}\right|, \quad \theta_1 \ = \ \left|\frac{\ln a_1}{h_1}\right|.$$

2. 本当のＢＰではない！
   - $\Pr[C_i = 1]$ と $\Pr[C_i = 0]$ の比の対数を求めることにした.
   - メッセージの伝播で大きな簡単化を 1 つ行った.

実際動かしていると非常にうまく働く. $\boxed{\text{何とか証明したい！}}$

そこで

- アルゴリズムの平均的解析 $\Longrightarrow$ Planted Solution Model
- MAXSTEP $= 2 \Leftarrow$ それでも結構よく働く

平均時計算量のモデル

$$\textbf{Planted Solution Model} \;=\; \boxed{\text{入力グラフを生成する手法}}$$

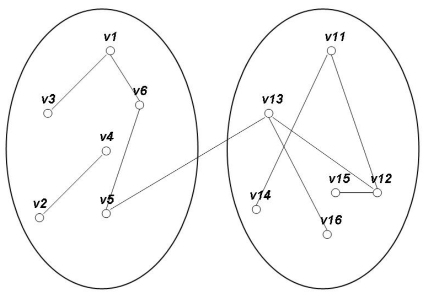
アルゴリズムの平均的解析のモデルとしてお勧め？

今回用いた Planted Solution Model

パラメータ n（頂点数 $2n$），$p, r, 0 < r < p < 1$ に対し，
次のようにグラフ $G = (V, E)$ を生成する．

1. $V_0 = \{1, ..., n\}$, $V_1 = \{n+1, ..., 2n\}$, and $V = V_0 \cup V_1$.
2. Put an edge (i, j) into E
   with probability p if $C_i = C_j$, and
   with probability r if $C_i \neq C_j$,

where $C_i = 0 \iff i \in V_0$ and $C_i = 1 \iff i \in V_1$.



要するに

答えを作り，それを元に問題を（ランダムに）作成する方法.

重要な性質

定理． パラメータ n, p, r の間に

$$p - r \;=\; \Omega\left(\frac{1}{\sqrt{n}}\right)$$

という関係があるとき，埋め込んだ解 (planted solution) が，生成したグラフに対する Bisection Problem の唯一の解となる．

$\Downarrow$

Planted Solution Model のもとでは，アルゴリズムの正当性は，生成したグラフに対し，埋め込んだ解が正しく見出せるか否かを調べればよい．

得られた結果

定理． Planted Solution Model のもとでアルゴリズムを実行させたとき,

$$\Pr[\text{ the algorithm yields the planted solution }] \;\geq\; 1 - 2n \cdot e^{-\epsilon_1 n \cdot \frac{(p-r)^4}{p^2}}.$$

系．

For any $\delta > 0$ and any p and r, $cp < r < p < 1$,
$p - r \geq c_2 n^{-1/2} \log(1/(\delta(p-r))))$
$\Rightarrow$ the algorithm answers correctly with probability $> 1 - \delta$.

これまでのアルゴリズムとの比較：

(Boppana, 1987)	$p - r = \Omega(n^{-1/2})$ (almost!?)
(Dyer and Frieze, 1989)	$p - r = \Omega(n^{-1/2})$ when p is large, e.g., $\Omega(1)$
(Jerrum and Sorkin, 1998)	$p - r = \Omega(n^{-1/6})$
(Condon and Karp, 2001)	$p - r = \Omega(n^{-1/2+\varepsilon})$

その他にも…

- Most Likely Partition 問題（with Onsjoe）
- MAX-2SAT 問題に対する同様のアプローチ（with 山本）
- 固有値を求める手法にできないか？

未解決問題：本当のＢＰの妥当性？

参考文献

R. McEliece, D. MacKay, and J.F. Cheng, Turbo decoding as an instance of Pearl's "Belief Propagation" algorithm, *IEEE J. on Selectead Areas in Communications*, 16(2):140–152, 1998.