

ランダムハイパーグラフの分割に対する カットサイズ分布

野崎 隆之

山口大学

情報理論研究会

2016/1/18

概要

これまでの研究

低密度パリティ検査 (LDPC) 符号に対する並列符号化法の提案
(手法) **ハイパーグラフの分割**によってパリティ検査行列をブロック対角化

本研究の目的

符号化法の計算時間の解析 \approx 並列数の解析

今回の成果

- 並列数はハイパーグラフのカットサイズに依存
- 二分割でのカットサイズ分布・典型最小カットサイズを正則 LDPC アンサンブルに対して導出

関連研究

- 符号理論的アプローチによるグラフの解析
[Fujii-Wadayama2012], [Yano-Wadayama2012], [Fujii-Wadayama2013]
- 情報統計力学的アプローチによるグラフの分割の解析 [Dembo et al.2015]

LDPC 符号

(定義) LDPC 符号

疎なパリティ検査行列 \mathbf{H} によって定義される符号

$$\mathcal{C} := \{x \in \mathbb{F}^n \mid \mathbf{H}x = \mathbf{0}\}$$

符号化

メッセージ m から符号語 $x = (p, m)$ を決定する操作

$$\begin{aligned}\mathbf{H}x &= \mathbf{0} \\ \iff (\mathbf{H}_P \quad \mathbf{H}_I) (p, m) &= \mathbf{0} \\ \iff \mathbf{H}_P p &= -\mathbf{H}_I m\end{aligned}$$

線形方程式を解きやすいような \mathbf{H}_P を行置換・列置換で作成

LDPC 符号の並列符号化法 [Nozaki2015]

パリティ検査行列を次のようにブロック対角化 ($\mathbf{H}_{P,i}$ は正則行列)

$$\mathbf{P}\mathbf{H}\mathbf{Q} = (\mathbf{H}_P \quad \mathbf{H}_I) = \begin{pmatrix} \mathbf{H}_{P,1} & & \mathbf{O} & \mathbf{H}_{I,1} \\ & \ddots & & \vdots \\ \mathbf{O} & & \mathbf{H}_{P,K} & \mathbf{H}_{I,K} \end{pmatrix},$$

符号化の手順

- $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K)$
- $\forall i \in [1, K] \quad \mathbf{H}_{P,i}\mathbf{p}_i = -\mathbf{H}_{I,i}\mathbf{m}$ を並列に解く

総計算量は既存法 [RU2001] とほぼ同じ

⇒ 並列化によって計算時間が既存法の $1/K$

ここからの流れ

- どうやってブロック対角化するか? (ハイパーグラフの分割)
- 並列数 K はどこまで大きく出来るか?
($K \geq 2$ になる必要条件を導出)

符号のハイパーグラフ表現

(定義) ハイパーグラフ $\mathcal{H} = (\mathcal{U}, \mathcal{E})$

- \mathcal{U} : ノードの集合
- \mathcal{E} : ネット (ハイパーエッジ) の集合

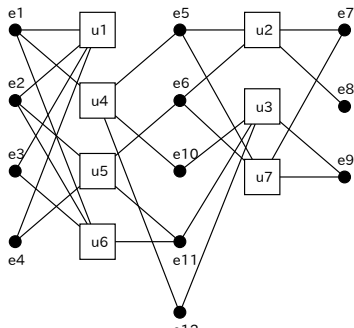
各ネットは1つ以上のノードに接続
(グラフ: 全てのネットが2つのノードに接続)

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

行列のハイパーグラフ表現 (Row-net model)
[Çatalyürek 1999]

第 i 列目 \Rightarrow 第 i ネット e_i

第 j 行目 \Rightarrow 第 j ノード u_j



ハイパーグラフの分割

ハイパーグラフの K 分割 $\Pi = \{\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_k\}$

(1) $\emptyset \neq \mathcal{U}_i \subseteq \mathcal{U}$, (2) $\mathcal{U}_i \cap \mathcal{U}_j = \emptyset$ (for $i \neq j$), (3) $\bigcup_{i=1}^K \mathcal{U}_i = \mathcal{U}$

カット：2つ以上の part に接続するネット

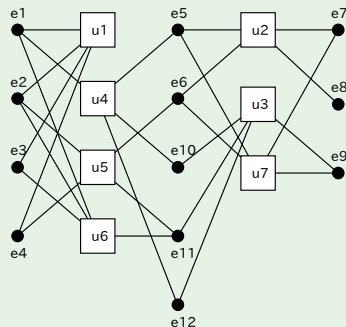
$\mathcal{X}(\Pi)$ ：分割 Π に対するカットの集合 (カットセット)

$$K = 2$$

$$\mathcal{U}_1 = \{u_1, u_4, u_5, u_6\}$$

$$\mathcal{U}_2 = \{u_2, u_3, u_7\}$$

$$\mathcal{X}(\Pi) = \{e_5, e_6, e_{10}, e_{11}, e_{12}\}$$



ハイパーグラフの分割と行列の構成

ハイパーグラフの K 分割問題

minimize : $|\mathcal{X}(\Pi)|$

s.t. ϵ -バランス条件 $\max_i |\mathcal{U}_i| < |\mathcal{U}|(1 + \epsilon)/K$

ヒューリスティックアルゴリズムが存在 (ex: PaToH [Çatalyürek])

行列のブロック化 ($K = 2$)

$$\mathbf{H} = \begin{matrix} & \begin{matrix} \text{(a)} & \text{(b)} & \text{(c)} \end{matrix} \\ \begin{matrix} \text{(1)} \\ \text{(2)} \end{matrix} & \begin{pmatrix} \mathbf{A}_1 & \mathbf{O} & \mathbf{B}_1 \\ \mathbf{O} & \mathbf{A}_2 & \mathbf{B}_2 \end{pmatrix} \end{matrix}$$

(1) \mathcal{U}_1
(2) \mathcal{U}_2
(a) \mathcal{U}_1 のみに接続するネット
(b) \mathcal{U}_2 のみに接続するネット
(c) **カットセット**

(命題) 符号 $\mathbf{H} \in \mathbb{F}^{m \times n}$ が K 並列符号化可能であるための必要条件

$$n - m \geq \min_{\Pi_K^{(\epsilon)}} |\mathcal{X}(\Pi_K^{(\epsilon)})|$$

グラフアンサンブル

(定義) 正則 LDPC アンサンブル $E(n, \gamma, \delta)$

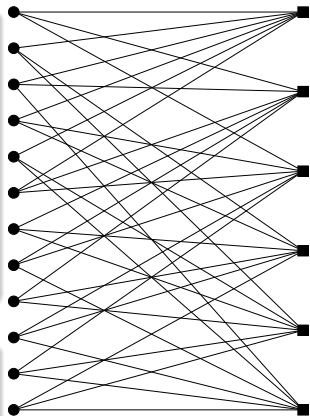
タナーグラフ $G = (V \cup C, E)$ の集合

- V : 変数ノードの集合
(次数 γ , 個数 n)
- C : 検査ノードの集合
(次数 δ , 個数 $m = \gamma n / \delta$)
- E : 辺の集合 (本数 $\xi = \gamma n$)

ハイパーグラフアンサンブル $E(n, \gamma, \delta)$

正則 LDPC アンサンブル $E(n, \gamma, \delta)$ を次のとおりに変形

- j 番目の変数ノード $\Rightarrow j$ 番目のネット
- i 番目の検査ノード $\Rightarrow i$ 番目のノード



2分割のカットサイズ分布 (定義)

$\mathcal{H} \in \mathbb{E}(n, \gamma, \delta)$ 2分割 $\Pi_2 = (\mathcal{U}_1, \mathcal{U}_2)$

■ $A_{\mathcal{H}}(s, m_1)$: \mathcal{H} に対して以下を満たす 2分割 Π_2 の個数

■ $s = |\mathcal{X}(\Pi_2)|$

■ $m_1 = |\mathcal{U}_1|$

■ $A(s, m_1) := \mathbb{E}_{\mathcal{H} \in \mathcal{E}}[A_{\mathcal{H}}(s, m_1)]$ (アンサンブル平均)

$$A(s, m_1) = [\sum_{\mathcal{C} \in \mathcal{E}} A_{\mathcal{C}}(s, m_1)] / \xi!$$

■ $B_{\mathcal{H}}(s, \epsilon)$: \mathcal{H} に対して以下を満たす 2分割 Π_2 の個数

■ $s = |\mathcal{X}(\Pi_2)|$

■ ϵ -balance 条件を満たす ($m_1 \in M_{\epsilon} := \llbracket m(1 - \epsilon)/2, m(1 + \epsilon)/2 \rrbracket$)

■ $B(s, m_1) := \mathbb{E}_{\mathcal{H} \in \mathcal{E}}[B_{\mathcal{H}}(s, m_1)]$ (アンサンブル平均)

$$B(s, \epsilon) = \sum_{m_1 \in M_{\epsilon}} A(s, m_1)$$

2分割のカットサイズ分布 (定理)

アンサンブル $E(n, \gamma, \delta)$ に対して, $A(s, m_1)$ は次式の通りに与えられる.

$$A(s, m_1) = \frac{\binom{m}{m_1} \binom{n}{s}}{\binom{\delta m}{\delta m_1}} \text{Coef}(f(u)^n, u^{\delta m_1}) \mathbb{I}[s \leq \delta m_1] \mathbb{I}[s \leq \delta(m - m_1)],$$

$$f(u) := p(u)^{s/n} q(u)^{1-s/n},$$

$$p(u) := (1 + u)^\gamma - 1 - u^\gamma, \quad q(u) := 1 + u^\gamma.$$

$\text{Coef}(f(x), x^i)$: 多項式 $f(x)$ の x^i の係数

$\mathbb{I}[P]$: P が真の時に 1 になり, P が偽のときに 0 (指示関数)

証明の概略: LDPC アンサンブルの重み分布の導出法と同様

アクティブエッジ: タナーグラフ G に対して, u_1 に接続するエッジ

カットに含まれるネットはアクティブエッジに 1 から $\gamma - 1$ 本接続する変数ノードに対応

漸近解析 (Growth rate の定義)

$$g(\sigma, \mu_1) := \lim_{n \rightarrow \infty} \frac{1}{n} \log A(\sigma n, \mu_1 n)$$

$$h(\sigma, \epsilon) := \lim_{n \rightarrow \infty} \frac{1}{n} \log B(\sigma n, \epsilon)$$

$$A(\sigma n, \mu_1 n) \approx 2^{ng(\sigma, \mu_1)}$$

$h(\sigma, \epsilon) < 0$ ならば n に対して指数関数的に減少

$h(\sigma, \epsilon) > 0$ ならば n に対して指数関数的に増加

⇒ 典型最小カットサイズ $\beta_\epsilon^* := \inf\{\sigma > 0 \mid h(\sigma, \epsilon) > 0\}$ が重要 (

$h(\beta_\epsilon^*, \epsilon) = 0$)

$$h(\sigma, \epsilon) = \max_{\mu_1 \in M_\epsilon} g(\sigma, \mu_1)$$

漸近解析 (Growth rate の導出)

$\sigma \leq \min\{\gamma\mu_1, \gamma(1 - \mu_1)\}$ の時, $E(\gamma, \delta)$ に対する Growth rate $g(\sigma, \mu_1)$, $h(\sigma)$ は次式で与えられる.

$$g(\sigma, \mu_1) = H_2(\sigma) - \gamma \frac{\delta - 1}{\delta} H_2(\mu_1) \\ + \inf_{u>0} \{ \sigma \log p(u) + (1 - \sigma) \log q(u) - \mu_1 \gamma \log u \}$$
$$h(\sigma) = \max_{\mu_1 \in M_\epsilon} g(\sigma, \mu_1)$$

極小値をとる u は次式で与えられる.

$$\sigma u p'(u) q(u) + (1 - \sigma) u p(u) q'(u) = \mu_1 \gamma p(u) q(u).$$

なお, $H_2(x) = -x \log x - (1 - x) \log(1 - x)$ (2元エントロピー関数)

漸近解析 (Growth rate の性質)

1 $h(\sigma, \epsilon)$ に対する閉じた下界

$$\begin{aligned} h(\sigma, \epsilon) &\geq h(\sigma, 0) = g(\sigma, 1/2) \\ &= H_2(\sigma) + \sigma \log(2^{\gamma-1} - 1) - \gamma \frac{\delta - 1}{\delta} + 1. \end{aligned}$$

2 $\sigma = 0$ のときは Growth rate が負 ($\gamma \geq 2, \delta \geq 3$)

$$g(0, \mu_1) = \left(1 - \gamma \frac{\delta - 1}{\delta}\right) H(\mu_1), \quad h(0, \epsilon) = 1 - \gamma \frac{\delta - 1}{\delta}$$

3 Growth rate が正になる σ が必ず存在

$\sigma^+ := 1 - (1 - \mu_1)^\gamma - \mu_1^\gamma \leq \min\{\gamma\mu_1, \gamma(1 - \mu_1)\}$ において

$$g(\sigma^+, \mu_1) = \frac{\gamma}{\delta} H_2(\mu_1) > 0.$$

4 2, 3 より典型最小カットサイズは必ず存在

漸近解析 (典型最小カットサイズの性質)

$$\beta^*(\epsilon) := \inf\{\sigma > 0 \mid h(\sigma, \epsilon) > 0\}$$

典型最小カットサイズの上界

$E(\gamma, \delta)$ に対して, 次の関係式が成り立つ.

$$\beta^*(\epsilon) \leq \beta^*(0).$$

注意: $h(\sigma, 0)$ は閉じた式で与えられるので $\beta^*(0)$ は容易に計算可能

典型最小カットサイズを利用した 2 並列に符号化可能であるための必要条件

$$1 - \delta/\gamma \geq \beta^*(\epsilon)$$

証明: $\sigma < \beta^*(\epsilon)$ ならば指数関数的に分割の個数が減少 (ほぼ存在しない)

$$n - m \geq \min_{\Pi_2^{(\epsilon)}} |\mathcal{X}(\Pi_2^{(\epsilon)})| \iff 1 - \delta/\gamma \geq \min_{\Pi_2^{(\epsilon)}} |\mathcal{X}(\Pi_2^{(\epsilon)})|/n$$

数値例 1

$\gamma = 2$ の場合

| δ | 3 | 4 | 5 | 6 | 7 | 8 |
|---------------------|--------|--------|--------|--------|--------|--------|
| $1 - \gamma/\delta$ | 0.3333 | 0.5000 | 0.6000 | 0.6667 | 0.7142 | 0.7500 |
| $\beta^*(0)$ | 0.0615 | 0.1100 | 0.1461 | 0.1740 | 0.1962 | 0.2145 |

(2, δ)-正則 LDPC 符号はいつでも並列符号化の必要条件を満たす

$\gamma = 3$ の場合

| δ | 4 | 5 | 6 | 7 | 8 | 9 |
|---------------------|--------|--------|--------|--------|--------|--------|
| $1 - \gamma/\delta$ | 0.2500 | 0.4000 | 0.5000 | 0.5714 | 0.6250 | 0.6667 |
| $\beta^*(0)$ | 0.2636 | 0.3157 | 0.3545 | 0.3849 | 0.4094 | 0.4297 |

(3, 4)-正則 LDPC 符号は符号化を並列にできない

(3, δ)-正則 LDPC 符号 ($\delta \geq 5$) は並列符号化の必要条件を満たす

(実際に (3,5)-正則 LDPC 符号は $n > 300$ 程度で並列符号化が可能)

数値例 2

$$\gamma = 5$$

| δ | 6 | 10 | 15 | 20 | 21 | 25 |
|---------------------|--------|--------|--------|--------|--------|--------|
| $1 - \gamma/\delta$ | 0.1667 | 0.5000 | 0.6667 | 0.7500 | 0.7619 | 0.8000 |
| $\beta^*(0)$ | 0.5570 | 0.6589 | 0.7193 | 0.7537 | 0.7589 | 0.7764 |

(5, δ)-正則 LDPC 符号 ($\delta \leq 20$) は符号化を並列にできない

(5, δ)-正則 LDPC 符号 ($\delta \geq 21$) は並列符号化の必要条件を満たす

実際にブロック対角化による符号化法を試してみると
 $\delta = 21, 22, 23, 24, 25$ のときは並列符号化ができない...

原因：

ハイパーグラフの分割問題は NP 困難

⇒ ヒューリスティックアルゴリズムを利用するしかない

⇒ カットサイズが最小となるような分割を得られていない?

まとめと今後の課題

まとめ

- K 並列符号化可能であるための必要条件を与えた
 - 符号 $\mathbf{H} \in \mathbb{F}^{m \times n}$ に対して,

$$n - m \geq \min_{\Pi_K^{(\epsilon)}} |\mathcal{X}(\Pi_K^{(\epsilon)})|$$

- アンサンブル中の符号が高い確率で 2 並列符号化可能

$$1 - \delta/\gamma \geq \beta^*(\epsilon)$$

- カットサイズ分布とその growth rate の導出
- 数値例による検証

今後の課題

- K 分割のカットサイズの導出
- 非正則 LDPC 符号への拡張