1　ハフマン符号のコンパクト性

1.1　コンパクト符号の符号木についての性質

Lemma 1. 根を除く全ての頂点が r 個の兄弟を持つ木 *(*全 r 分木という *)* は $1+k(r-1)$ 個の葉を持つ*.*

Proof. 子供を持つ頂点の数に関する帰納法. □

Lemma 2. φ を情報源 (Σ, P) の r 元コンパクト符号*,* (V, E) をその符号木とすると次を満たす*:*

1. $|\varphi(x)| < |\varphi(y)| \Rightarrow p_x \geq p_y$

2. 兄弟が自身を含めて r より真に少ない様な頂点は根からの最遠点である*.*

Proof.

1. x, y の符号語を入れ替えた符号 $\tilde{\varphi}$ の平均符号長は

$$|\tilde{\varphi}| = |\varphi| - (|\varphi(y)| - |\varphi(x)|)(p_y - p_x)$$

である. 最小性より $|\tilde{\varphi}| \geq |\varphi|$ であるので $p_x \geq p_y$.

2. $w\hat{a}$ を「兄弟が自身を含めて r より真に少ない様な頂点」, 即ち $\hat{a} \neq a \in A$ で wa が符号や他の符号語の接頭語で無いもの (即ち符号木の頂点でない) が存在すると仮定する.

wa 自身は符号語でなく, wa 自身以外の wa の接頭語は $w\hat{a}$ の接頭語でもあるので φ の瞬時復号化可能性から他の φ の符号語と一致しない. よって, ある $\widetilde{\sigma}$ の符号語を wa で置き換えた符号

$$\varphi_{\widetilde{\sigma}}(\sigma) := \begin{cases} wa & \text{if } \sigma = \widetilde{\sigma} \\ \varphi(\sigma) & \text{otherwise} \end{cases}$$

はやはり瞬時復号化可能性を持つ.

更に

$$|\varphi_{\widetilde{\sigma}}| = |\varphi| - (|\varphi(\widetilde{\sigma})| - |wa|)p_{\widetilde{\sigma}} = |\varphi| - (|\varphi(\widetilde{\sigma})| - |w\hat{a}|)p_{\widetilde{\sigma}}$$

であるが, 最小性より $|\varphi_{\widetilde{\sigma}}| \geq |\varphi|$ であるので $|\varphi(\widetilde{\sigma})| \leq |w\hat{a}|$. $\widetilde{\sigma}$ の任意性より $w\hat{a}$ は根からの最遠点.

□

$m := \#\Sigma$ 個の元を持つ情報源 $(\Sigma, (p_\sigma))$ に対し, $m-1$ を $r-1$ で割った商を切り上げたものを n とする；

$$m - 1 = (n-1)(r-1) + R, \quad (0 < R \leq r-1)$$

Proposition 1. 確率が小さい順に選んだ $d = R+1$ 個の元 $\{\sigma_1, \ldots, \sigma_d\}$ の符号語が兄弟であるようなコンパクト符合が存在する*.*

Proof. コンパクト符号は少なくとも一つ存在する. 任意のコンパクト符号から, コンパクト性を保つ変更を加えて構成する.

STEP1. コンパクト符号において, $\sigma \notin \{\sigma_1, \ldots, \sigma_d\}$ で $|\varphi(\sigma)| > |\varphi(\sigma_i)|$ なるものがあれば Lemma 2 から $p_\sigma \leq p_{\sigma_i}$ であって, 最小性より $p_\sigma = p_{\sigma_i}$ であるので, σ と σ_i の符号語を入れ替えた符号もまた同じ平均符号長で, 即ちコンパクト符号である. 必要なだけこの操作を繰り返せば, 任意の $\sigma \notin \{\sigma_1, \ldots, \sigma_d\}$ について $|\varphi(\sigma)| \leq |\varphi(\sigma_i)|$ なるコンパクト符号が得られる.

STEP2. 次に前段の操作でできたコンパクト符号は $|\varphi(\sigma_i)|$ が全て同じ長さである事を背理法で示す. $|\varphi(\sigma_i)| < \max_j |\varphi(\sigma_j)|$ なる i がある場合, 即ち最長の符号の数 q が d より真に小さいと仮定する. 符号木において最遠の葉を除いた木はコンパクト性より Lemma 2 から全分木であるが, その葉の数 l は

$$m - (q-1) \leq l < m$$

である, ここで $q - 1 < R$ を使って

$$1 + (n-1)(r-1) < 1 + (n-1)(r-1) + R - (q-1) = m - (q-1) \leq l < m \leq 1 + n(r-1)$$

これは Lemma 1 から全分木であることに矛盾する.

STEP3. 同じ距離であればどの葉を用いても, 瞬時符号化性と平均符号長を保つので, $\{\sigma_1, \ldots, \sigma_d\}$ の符号語が同じ親を持つようにコンパクト性を保ちつつ符号を修正できる. ($d < r$ である場合も全分でない頂点は全て最遠であるから, 長さを変えず d 個の兄弟を持つ符号語が作れる)

□

2 コンパクト性の遺伝

Lemma 3. 情報源 $(\Sigma', (p'_{\sigma'}))$ が情報源 $(\Sigma, (p_\sigma))$ の縮退化情報源で,

$$1 + k(r-1) = \#\Sigma' < \#\Sigma = 1 + k(r-1) + R \quad (1 < R \leq r-1)$$

なるものであり, 次を満たす縮退化写像 $\iota : \Sigma \to \Sigma'$ が存在するとする*;* $\hat{\sigma}' \in \Sigma'$, $\{\sigma_1, \ldots, \sigma_d\} \subset \Sigma$*(*ただし $d = R + 1$*)* が存在して

1. $\iota(\sigma_i) = \hat{\sigma}'$ *for* $i = 1, 2, \ldots, d$

2. $\iota|_{\Sigma \setminus \{\sigma_1, \ldots, \sigma_d\}}$ は単射.

3. $\{\sigma_1, \ldots, \sigma_d\}$ は確率最小のもの. *i.e.,* $p_{\sigma_i} \leq p_\sigma$ *for any* i *and* $\sigma \in \Sigma \setminus \{\sigma_1, \ldots, \sigma_d\}$

このとき情報源 $(\Sigma', (p'_{\sigma'}))$ の r 元コンパクト符号 φ' と $\{a_1, \ldots, a_d\} \subset A$ から

$$\varphi(\sigma) = \begin{cases} \varphi'(\hat{\sigma}')a_i & \textit{if } \sigma = \sigma_i \\ \varphi'(\iota(\sigma)) & \textit{otherwise} \end{cases}$$

で与えられる符号 φ は情報源 $(\Sigma, (p_\sigma))$ の r 元コンパクト符号となる.

Proof. 定義から φ も瞬時復号化可能で $|\varphi| = |\varphi'| + \sum_{i=1}^{d} p_{\sigma_i}$ である.

Proposition 1 から Σ のコンパクト符号で $\{\sigma_1, \ldots, \sigma_d\}$ の符号語が同じ親 w を持つもの $\hat{\varphi}$ が存在する. この $\hat{\varphi}$ から

$$\hat{\varphi}'(\sigma) = \begin{cases} w & \text{if } \sigma = \sigma_i \\ \varphi(\sigma) & \text{otherwise} \end{cases}$$

で定義される $\hat{\varphi}'$ は Σ' の瞬時復号化可能かつ $|\hat{\varphi}'| = |\hat{\varphi}| - \sum_{i=1}^{d} p_{\sigma_i}$. φ' のコンパクト性から

$$|\hat{\varphi}| = |\hat{\varphi}'| + \sum_{i=1}^{d} p_{\sigma_i} \geq |\varphi'| + \sum_{i=1}^{d} p_{\sigma_i} = |\varphi|$$

よって φ もコンパクト. □