

ρセクタ問題およびデータ変換に関連した問題の計算複雑さ

広島大学 工学部

菊野 亨

角田 良明

若林 真一

杉原 一夫

1. ρセクタ問題に関連した問題の計算複雑さ

菊野, 角田

1.1 まえがき

コミュニケーションネットワークにおける中継局の最適配置, 都市計画における公共施設(例えば, 病院)の最適配置に関する基礎的研究として, ρセクタ問題はよく知られている.⁽¹⁾⁽³⁾ 最近, ρセクタ問題に幾つかの制限を付けた問題について, 特に時間計算量に関する多くの報告がなされてきており, Kariv & Hakimi は制限されたρセクタ問題がNP-完全であることを示している.⁽²⁾

本論文では先ず次のρセクタ問題を定義し, これがやはりNP-完全であることを示す.

条件付pセンタ問題

入力：無向グラフ $G = (V, E)$ ， V の真部分集合 S ，2つの正整数 k, ℓ

判定：次の条件を満たす V の部分集合 D (以降この集合 D を S に関する G の pセンタ集合と呼ぶ) で， $|D| \leq k$ なるものが存在するか否か？

条件：節点 $v \in S$ と D 内の節点との間の最短道の長さを $d(v)$ とかく時， $\max_{v \in S} d(v) \leq \ell$ が成立する。

次に，条件付pセンタ問題の1つの応用として，複合交換方式のコンピュータネットワークにおける中継局の最小化問題⁽⁴⁾がNP-完全となることも証明できるが省略する。

1.2 NP-完全性

無向グラフ $G = (V, E)$ と非負整数 d に対し，各節点 $v \in V$ はひとの距離が d 以下の全ての節点を "距離 d で支配する" という。

定理 1.1 条件付pセンタ問題はNP-完全である。

(証明) 条件付pセンタ問題がクラスNPに属することは明らかなので，以下では $G = (V, E)$ と2つの正整数 k, ℓ に

対する制限付 ρ センタ問題[†]からの多項式時間の変換が存在することを示す. G と ρ に基づき, 無向グラフ $G' = (V', E')$ と V' の真部分集合 S を次の様に構成する (図 1.1 参照). $V = \{v_1, v_2, \dots, v_n\}$ とする.

$$V' = V \cup \{v_i^j \mid v_i \in V, 1 \leq j \leq 2\rho + 1\}$$

$$E' = E \cup \{(v_i^j, v_i^{j+1}) \mid v_i \in V, 0 \leq j \leq 2\rho\}$$

但し, $v_i = v_i^0$ とする.

$$S = V \cup \{v_i^{2\rho+1} \mid v_i \in V\}$$

更に, 正整数 $k' = k + n$ とおく.

G の ρ センタ集合 D で, $|D| \leq k$ なるものが存在するための必要十分条件は S に関する G' の ρ センタ集合 D' で, $|D'| \leq k'$ なるものが存在することを以下で示す.

必要性: $D' = D \cup \{v_i^{2\rho+1} \mid v_i \in V\}$ と決める. D が G の ρ センタ集合であるので, G' 上で (V' の部分集合) V に属するすべての節点は D' の部分集合 D に属するいずれかの節点到距離 ρ で支配されている. 更に, 各 v_i^j ($1 \leq j \leq 2\rho + 1$) は $v_i^{2\rho+1}$ に距離 ρ で支配される. よって, D' は S に関する G' の ρ センタ集合である. D' の構成法より $|D'| = |D| + n \leq k + n = k'$ が

[†] 条件付 ρ センタ問題において $S = V$ を仮定すると 制限付 ρ センタ問題となる. 従って制限付 ρ センタ問題では D を単に G の ρ センタ集合と呼ぶことにする.

成立する。

十分性: $D = \{v_i \mid \text{少なくとも1つの } j (0 \leq j \leq c) \text{ に対し } v_i^j \in D' \text{ が成立}\}$ と決める。 G' 上で S に属する節点はすべて D に属するいずれかの節点到距離 c で支配されている。従って、各 v_i^0 はある v_i^j (必ずしも $i = i'$ とは限らぬ) に距離 c で支配される。今、 v_i^j の代わりに v_i^0 を考えても、各 v_i^0 が距離 c で支配されることに変化はない。よって D は G の ρ センタ集合である。

今、 $A = D' \cap \{v_i^j \mid 0 \leq j \leq c\}$ とすると、 $|A| = |D| + \alpha$ が成立する。ここで α はある非負整数で、 G' 上で S に属する節点で特に V に属する節点到距離 c で支配している節点の内、冗長なもの数に対応している。更に各 v_i^{2c+1} ($1 \leq i \leq n$) を支配するためには $c+1 \leq j \leq 2c+1$ なる少なくとも1つの j に対し、 v_i^j が D' に属しているはずである。しかもこの v_i^j は決してどの v_i^0 ($1 \leq i \leq n$) も距離 c で支配することはないので、こうした v_i^j の総数は $n + \beta$ となる。ここで β はある非負整数で、やはり冗長な節点の総数に対応する。以上より $|D'| = |D| + n + \alpha + \beta$ が成立する。 $|D'| \leq k' = n + k$ より、 $|D| \leq k - (\alpha + \beta) \leq k$ となる。(証明終)

なお、残された問題としては条件付 ρ センタ問題に対する NP-完全性と、 G の形状及び集合 S との関連を明らかにす

ることがあり，現在検討中である．

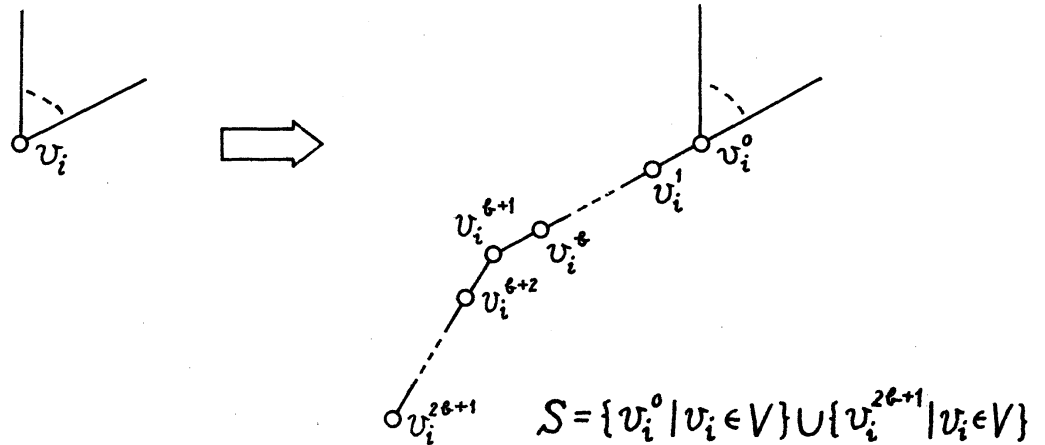


図1.1 定理1.1の変換

謝辞

文献を御提供して下さい下さった京都大学茨木俊秀助教授並びに御討論頂いた大阪大学谷口健一助教授に深謝いたします。

文献

- (1) Christofides, N. : "Graph theory : an algorithmic approach", Academic Press (1975).
- (2) Garey, M.R. and Johnson, D.S. : "Computers and intractability", W. H. Freeman and Company (1979).
- (3) T. Ibaraki, S. Naito and T. Hasegawa : "The complexity of p-center problems on networks", Working Paper, Dept. of Applied Mathematics and Physics, Kyoto University (1979).
- (4) 菊野, 吉田, 角田, 原田 : "ネットワークにおける通過節点数の最小化", 信学技報, CAS79-154 (1980).

2. コモンスーパーシーケンスに関する問題の計算複雑さ

菊野, 若林

2.1 まえがき

プログラムテキストの編集およびアミノ酸系列の識別に関する基礎的研究として, 系列の集合に対する2, 3の問題に興味をもたれてきている⁽¹⁾⁽²⁾ アルファベット Σ 上の有限系列 $w = \sigma_1 \sigma_2 \dots \sigma_m$ (各 $\sigma_i \in \Sigma$) に対し w の任意の要素を削除して得られる系列を $w' = \sigma_{i_1} \sigma_{i_2} \dots \sigma_{i_n}$ とする (従って逆に, w' に対し Σ に属する要素を挿入すれば w が得られる). このとき w と w' の間に 関係 $w > w'$ が存在すると定義し, w を w' の スーパーシーケンス といい, 逆に w' を w の サブシーケンス という. Σ 上の有限系列の有限集合 $R = \{w_1, w_2, \dots, w_p\}$ ($p \geq 2$) に対し, $\hat{w} > w_i$ ($1 \leq i \leq p$) なる系列 \hat{w} を R の コモンスーパーシーケンス (cs 系列) と定義する. R の cs 系列の集合を $CS(R)$ で表わす.

集合 $CS(R)$ に関する判定問題,

SCS問題: Σ 上の系列の集合 $R = \{w_1, w_2, \dots, w_p\}$ と非負整数 k に対し, $w \in CS(R)$ かつ $|w| \leq k$ なる系列 $w \in \Sigma^*$ が存在するか否か,

について次の様な多くの結果が得られている.

F1⁽³⁾ $p = 2$ のとき SCS 問題は多項式時間で解ける

(更に系列 w も求まる).

F2⁽⁴⁾ $p \geq 3$ かつ $|\Sigma| = 3$ ぞ SCS問題 は NP-完全となる.

F3 (Rに含まれる) 系列の長さを制限した問題⁽⁵⁾ は NP-完全となる.

本論文では Rに含まれる系列は全マプリミティブだと更に制限しても(制限の詳細は2.2で述べる) SCS問題は NP-完全となることを示す.

2.2 判定問題

Σ 上の系列が Σ の中のどの文字も高々1つしか含んでいないとき系列 w はプリミティブであると定義する. プリミティブな系列の集合に対する SCS問題(以降では単にプリミティブな SCS問題と呼ぶ): $R = \{w_1, w_2, \dots, w_p\}$ ($p \geq 3$) を Σ 上の系列の有限集合とする. ここで各 w_i はプリミティブであって, 2つの系列 w_{p-1}, w_p の長さは任意の有限な値をとれるが残りの系列 w_i ($1 \leq i \leq p-2$)はいずれも長さが5の系列とする. このとき非負整数 k に対し, $w \in CS(R)$ かつ $|w| \leq k$ なる系列 $w \in \Sigma^*$ が存在するか否か.

2.3 NP-完全性

ここでは文献(1), (4)~(6)と同様, 既にNP-完全であることが示されている無向グラフ $G = (V, E)$ のNode Cover問題を利用してプリミティブなSCS問題がやはりNP-完全であるこ

とを示す。以降では $|V|=t, |E|=r$ とし, 更に V および E の各要素には $v_i (1 \leq i \leq t), e_i (1 \leq i \leq r)$ でそれぞれラベル付けし, そのラベルで各節点および枝を表わす。従って枝 e_i が2つの節点 v_j と v_m を結ぶものならば $e_i = (v_j, v_m)$ が成り立つものとする。

Node Cover問題とは「正整数 k と無向グラフ $G=(V, E)$ 上の枝のリスト $R; e_1; e_2; \dots; e_r$ が与えられているとき, $|N_0| = k$ かつ各枝 $e_i = (v_j, v_m)$ に対し $v_j \in N_0, v_m \in N_0$ の少なくとも1つが成立する V の部分集合 N_0 が存在するか否か」を判定する問題である。なお証明の便宜上, 各枝 $e_i = (v_j, v_m)$ に対し, $j < m$ が常に成立するものと仮定し, その枝を v_j から v_m への有向枝と見なす。従ってグラフ G には自己ループは存在しないものとする。

2.3.1 R の構成

無向グラフ $G=(V, E)$, 3つの定数 k, t, r は先に述べた通りとする。アルファベットを

$$\Sigma = \{v_i | 1 \leq i \leq t\} \cup \{e_i, \bar{e}_i, \hat{e}_i | 1 \leq i \leq r\} \cup \{b_i | 1 \leq i \leq 3c\}$$

と決める。但し $c = \max(r, t)$ とする。 Σ 上のプリミティブな系列の集合を

$$R = \{T_L, T_R\} \cup \{w_i | 1 \leq i \leq r\}$$

と定義する。ここで T_L, T_R , 各 w_i は次の通りとする。

1) $T_L = N \cdot E \cdot \bar{E} \cdot B$, $T_R = B \cdot \hat{E} \cdot E \cdot N$ とする. 但し

$$N = v_t v_{t-1} \cdots v_1, E = e_1 e_2 \cdots e_r, \bar{E} = \bar{e}_1 \bar{e}_2 \cdots \bar{e}_r,$$

$$\hat{E} = \hat{e}_1 \hat{e}_2 \cdots \hat{e}_r, B = b_1 b_2 \cdots b_{3c}$$

とする.

2) 各枝 $e_i = (v_j, v_m) \in E$ に対し $w_i = \bar{e}_i v_j e_i v_m \hat{e}_i$ とする.

[注1] T_L と T_R を連結して得られる系列 $N \cdot E \cdot \bar{E} \cdot B \cdot B \cdot \hat{E} \cdot E \cdot N$ から, 中央部に含まれる系列 $B \cdot B$ を削除すると文献(5)で構成した系列 T に等しくなる.

[注2] R に属する系列はいずれもプリミティブであって, その長さは $|T_L| = |T_R| = 2r + t + 3c$ (有限), $|w_i| = 5$ である.

補題1 無向グラフ $G = (V, E)$ ($t = |V|, r = |E|$) 上で $|N_0| = k$ なる Node Cover N_0 が存在するなら, 上述の通り構成したプリミティブな系列の集合 R に対しその長さが $5r + 2t + 3c + k$ の c s 系列が存在する.

(略証) G 上の Node Cover N_0 に基づいて Σ の部分集合 W, U を次の様に定義する.

$$W = \{ \hat{e}_i \mid e_i = (v_j, v_m) \in E \text{ かつ } v_j \in N_0 \}$$

$$U = \{ \bar{e}_i \mid e_i \in E \text{ かつ } \hat{e}_i \notin W \}$$

ここで $|W| = s, |U| = p$ とする (W, U の定義より $s + p = r$ となる). 次に3つの系列 $\hat{E}_W, \bar{E}_U, N_c$ をそれぞれ系列 \hat{E}, \bar{E}, N のサブシーケンスとして以下の様に定義する.

$$\hat{E}_w = \hat{e}_{i_1} \hat{e}_{i_2} \cdots \hat{e}_{i_s} \quad (\text{各 } \hat{e}_{i_q} \in W)$$

$$\bar{E}_u = \bar{e}_{j_1} \bar{e}_{j_2} \cdots \bar{e}_{j_p} \quad (\text{各 } \bar{e}_{j_q} \in U)$$

$$N_c = v_{i_1} v_{i_2} \cdots v_{i_r} \quad (\text{各 } v_{i_q} \in N_0)$$

今、次の様に定義される系列 T' を考えてみる。

$$T' = \bar{E}_u \cdot N \cdot E \cdot \bar{E} \cdot N_c \cdot B \cdot \hat{E} \cdot E \cdot N \cdot \hat{E}_w$$

すると、まず $T' > T_L$, $T' > T_R$ が成立することは明らか。次に Node Cover の定義より、各 $e_i = (v_j, v_m)$ に対し $v_j \in N_0$, $v_m \in N_0$ の少なくとも一方は成立している。そこで今 $v_j \in N_0$ とすると $\hat{e}_i \in W$ となり、 \bar{e}_i は系列 \bar{E} に、 v_j は系列 N_c に、 e_i は系列 E に、 v_m は系列 N に、 \hat{e}_i は系列 \hat{E}_w にそれぞれ含まれる。従って $T' > w_i$ が成立する。 $v_m \in N_0$ についても同様の議論により $T' > w_i$ が証明できる。よって T' は R の c s 系列となり、最後に T' の長さは $|T'| = |T_L| + |T_R| - |B| + |\bar{E}_u| + |\hat{E}_w| + |N_c| = 2(2r+t+3c) - 3c + p + s + r = 5r + 2t + 3c + r$ となる。 (略証終)

2.3.2 標準 c s 系列集合

2.3.1 で構成した R の c s 系列 w が次の条件 c を満たすならば、特に 標準 c s 系列 という。 R の標準 c s 系列の集合を $\underline{C} S_0(R)$ で表わす。

条件 c : 系列 w は、 $w = \alpha_1 \cdot N \cdot E \cdot \bar{E} \cdot \beta \cdot B \cdot \hat{E} \cdot E \cdot N \cdot \alpha_2$ である。但し、 $\alpha_1, \alpha_2, \beta$ は次の様に定義される系列とする。すなわち、今 N' を V の任意の部分集合とする (N' は必ずしも G の

Node Cover ではない). N' に基づいてまず Σ の部分集合 Σ_1, Σ_2 を, $\Sigma_1 = \{\hat{e}_i \mid e_i = (v_j, v_m) \in E \text{ かつ } v_j \in N'\}$, $\Sigma_2 = \{\bar{e}_i \mid e_i \in E \text{ かつ } \hat{e}_i \notin \Sigma_1\}$ と決める. 次に3つの系列 $\alpha_1, \alpha_2, \beta$ をそれぞれ \bar{E}, \hat{E}, N のサブシーケンスとして, $\alpha_1 = \bar{e}_{i_1} \bar{e}_{i_2} \dots \bar{e}_{i_p}$ (各 $\bar{e}_{i_q} \in \Sigma_2$), $\alpha_2 = \hat{e}_{j_1} \hat{e}_{j_2} \dots \hat{e}_{j_s}$ (各 $\hat{e}_{j_q} \in \Sigma_1$), $\beta = v_{i_1} v_{i_2} \dots v_{i_k}$ (各 $v_{i_q} \in N'$) と定義する.

[注3] 系列 w の中央部に系列 B が含まれている点を除けば, 条件 C は文献(5)の標準 CS 系列における w に対する定義と全く同じである.

このとき次の補題2が成立する. 証明は注1と注3で述べた事に注意すれば文献(5)の付録(補題3の証明)とほぼ同様に示せる.

補題2 2.3.1で構成したプリミティブな系列の集合 R に対し, $CS(R)$ を定義域, $CS_0(R)$ を値域とし, 次の条件1), 2)を満たす写像 φ が存在する.

- 1) φ は全射であって,
- 2) $w \in CS(R)$ に対し, $|w| \geq |\varphi(w)| > 5r + 2t + 3c$ が成立する.

補題3 2.3.1で構成したプリミティブな系列の集合 R に対し, 長さが $5r + 2t + 3c + k$ の CS 系列 w が存在するなら, 無向グラフ G 上で $|N_0| \leq k$ なる Node Cover N_0 が存在する.

(略証) 補題2より, c s 系列 w に対し $\varphi(w) \in CS_0(R)$ かつ $5r+2t+3c+k \geq |\varphi(w)| > 5r+2t+3c$ なる系列 $\varphi(w)$ ($=w'$ とおく) が存在する. 今 $|w'| = 5r+2t+3c+k'$ とすると $k \geq k' > 0$ となる. このとき集合 $CS_0(R)$ の定義の条件 c に現われた (系列 w のサブシーケンス) β に対応する系列 w' のサブシーケンス β は, 長さ k' の系列 $\beta = v_{i_1} v_{i_2} \dots v_{i_{k'}}$ である.

一方 R の定義より $w_i = \bar{e}_i v_j e_i v_m \hat{e}_i$ である. w が R の c s 系列であること, および φ の構成法より, w_i に現われる2つの文字 v_j と v_m の少なくとも1つは系列 $\beta = v_{i_1} v_{i_2} \dots v_{i_{k'}}$ の中に必ず含まれる. これが各 w_i について成立する.

故にこの β に基づいて $G = (V, E)$ の節点の集合 $N_0 = \{v_{i_1}, v_{i_2}, \dots, v_{i_{k'}}\}$ をとれば, この N_0 は G の Node Cover となる.

(略証終)

補題1, 3 および注2より次の定理が導ける.

定理 プリミティブな SCS 問題は NP -完全である.

2.4 おすび

これまで得られている結果をまとめると表2.1のようになる.

謝辞

有益な御討論を頂いた大阪大学谷口健一助教授に深謝いたします.

	コモンサブシーケンス に関する判定問題	コモンスーパーシーケンス に関する判定問題
一般	NP-完全	NP-完全
Σ に対する制限	$ \Sigma =2$ で NP-完全	$ \Sigma =3$ で NP-完全
$ w_i $ に対する制限	多項式時間	$ w_i =5 (1 \leq i \leq p-1)$ で NP-完全
プリミティブ	多項式時間	$ w_i =5 (1 \leq i \leq p-2)$ で NP-完全

表 2.1 コモンシーケンスに関する判定問題の計算複雑さ
文献

- (1) Maier, D.: "The complexity of some problems on subsequences and supersequences", JACM 25, 2, pp. 322-336 (1978).
- (2) Wagner, R. A. and Fisher, M. J.: "The string-to-string correction problem", JACM 21, 1, pp. 168-173 (1974).
- (3) Garey, M. R. and Johnson, D. S.: "Computers and Intractability", W. H. Freeman and Company (1979).
- (4) 菊野, 角田, 若林: "Supersequence に関する問題の計算複雑さ", 夏の LA シンポジウム資料 (1979).
- (5) 菊野, 吉田, 若林: "Shortest common supersequence に関する問題の計算複雑さ", 信学技報, AL 79-96 (1980).
- (6) 若林真一: "Supersequence に関する問題の計算複雑さ", ECS Lab. Hiroshima Univ. Tech. Rep. No. 79-07, 79-09, 79-14 (1979).

3. 木のエンコーディングにおけるホストグラフの計算量に関する比較

菊野, 杉原

3.1 まえがき

効率的なプログラムの開発において論理的なデータ構造を物理的な記憶構造に変換することは重要な操作の一つである。最近, この操作を一般的に, 且つ, 形式的に扱う研究, すなわち データエンコーディング あるいはデータの 記憶表現 について多くの報告がある⁽¹⁾⁻⁽³⁾。

通常, 論理的なデータ構造, 物理的な記憶構造をそれぞれ無向グラフ G, H で表し, G を ゲストグラフ, H を ホストグラフ と呼ぶ。但し, G を任意のグラフ, H を直線グラフとするとき, 変換コストがある定数以下となる G から H へのエンコーディングが存在するか否かの決定問題は NP-完全となることが既に示されている⁽²⁾。そこで以降では G を完全二分木に固定して議論する。

本論文では H としてデータ構造 DREE⁽¹⁾, TAP1⁽⁴⁾ および新たに提案する TAP3 を用いた場合の変換コストの比較を行う。更に G 上での探索を対応するデータ構造 H 上でシミュレートした場合の時間計算量の比較も行う。

3.2 準備

高さ h の 完全二分木 を $T(h) = (V, E)$ で表す. $T(h)$ の各節点には次の①, ②を満たす様, アルファベット $\Sigma = \{1, 2\}$ 上の系列がラベルとして付いている (図 3.1 参照). ①頂点のラベルは λ (空系列) である. ②節点 v のラベルを $\alpha \in \Sigma^*$ とすると, v の左子節点のラベルは $\alpha \cdot 1$, 右子節点のラベルは $\alpha \cdot 2$ である. 節点 v のラベルを $l(v)$ で表す.

定義 1⁽¹⁾ $T(h) = (V, E)$ に基づいて定義される無向グラフ $G = (V, E')$ を高さ h の DREE と呼び $D(h)$ で表す. ここで, $E' = E \cup \{(x, x') \mid x, x' \in V, l(x) \in \Sigma^{h-a-1}, a \geq 1, l(x') = l(x) \cdot 1 \cdot 2^a\} \cup \{(y, y') \mid y, y' \in V, l(y) \in \Sigma^{h-b-1}, b \geq 1, l(y') = l(y) \cdot 2 \cdot 1^b\}$ とする (図 3.2 参照).

定義 2 $T(h) = (V, E)$ に基づいて定義される無向グラフ $G = (V, E')$ を高さ h の TAP3 と呼び $T_3(h)$ で表す. ここで, $E' = E \cup \{(x, x') \mid x, x' \in V, l(x) = 1 \cdot \alpha \cdot 1 \cdot 2^a \in \Sigma^{h-1}, \alpha \in \Sigma^*, a \geq 0, l(x') = \sigma_h(1 \cdot \alpha)\} \cup \{(y, y') \mid y, y' \in V, l(y) = 2 \cdot \alpha \cdot 2 \cdot 1^b \in \Sigma^{h-1}, \alpha \in \Sigma^*, b \geq 0, l(y') = \sigma_h(2 \cdot \alpha)\} \cup \{(z, z') \mid z, z' \in V, l(z) = 1 \cdot 2^{h-2}, l(z') = 2 \cdot 1^{h-2}\}$ とする. 但し,

$$\sigma_h(\alpha) = \begin{cases} 1 \cdot \beta \cdot 1^{|\gamma|} \cdot 2^{h-|\beta|-|\gamma|-2} & (\alpha = 1 \cdot \beta \cdot \gamma, \beta \in \Sigma^*, \gamma \in \Sigma \text{ or } \beta = \gamma = \lambda) \\ 2 \cdot \beta \cdot 2^{|\gamma|} \cdot 1^{h-|\beta|-|\gamma|-2} & (\alpha = 2 \cdot \beta \cdot \gamma, \beta \in \Sigma^*, \gamma \in \Sigma \text{ or } \beta = \gamma = \lambda) \end{cases}$$

である (図 3.3 参照).

定義 1, 2 で与えたデータ構造 $D(h)$, $T_3(h)$ および $T(h)$ を特に区別する必要のない時は, それらを高さ h のデータ構造 $\hat{T}(h)$

として引用することにする。 $\hat{T}(h)$ と完全二分木 $T(h)$ との類似点を強調するため、 $\hat{T}(h)$ 内の節点が長さ h のラベルが付いているものを全て $\hat{T}(h)$ の葉と呼ぶことにする。

定義3 高さ h の完全二分木 $T(h) = (V, E)$ から高さ $h+1$ のデータ構造 $\hat{T}(h+1)$ への エンコーディング とは単射 $n: V \rightarrow L$ (L は $\hat{T}(h+1)$ の葉の集合) である。

定義4 定義3で与えた写像 n に基づいて単射 $\varepsilon: E \rightarrow P$ (P は $\hat{T}(h+1)$ 上の道の集合) を定義する。以降では、写像 ε のことも $T(h)$ から $\hat{T}(h+1)$ へのエンコーディングと呼ぶ。

定義5 $T(h) = (V, E)$ の E 上で定義される確率関数 $\pi: E \rightarrow [0, 1]$ を $T(h)$ の 利用頻度 という。 $T(h)$ から $\hat{T}(h+1)$ へのエンコーディング ε の利用頻度 π に関する 変換コスト は次式で与えられる。但し、 $lg(p)$ は道 p の長さを表す。

$$C_1(\varepsilon; \pi) = \sum_{e \in E} \pi(e) \cdot lg(\varepsilon(e))$$

定義6 $T(h)$ から $\hat{T}(h+1)$ へのエンコーディング ε の 演算コスト は次式で与えられる。

$$C_2(\varepsilon; i) = a \cdot A + t \cdot T + s \cdot S$$

但し、 $i = 1$ --- プレオーダー, 2 --- インオーダー, 3 --- ポストオーダー, 4 --- レベルオーダー,

A --- 節点へのアクセス操作

T --- ポインタ値の判定操作

S --- スタックあるいはキューのアクセス操作

a, t, s, \dots, i で指定する $T(h)$ 上の演算を $\hat{T}(h+1)$ 上でシミュレートするとき生じる上述の A, T, S の各操作の総数とする。

3.3 完全2分木のエンコーディング

ここでは高さ h の完全2分木 $T(h)$ から高さ $h+1$ の DREE および TAP3 へのエンコーディングの内、後で必要なものだけ例を用いて述べる。

エンコーディング $\varepsilon_h^{(1)}$ 形式的な定義は省き、DREE へのエンコーディング例だけ示す。 ε_3 は図 3.1 の $T(3)$ に対し、図 3.2 に示す通り、 $T(3)$ 上をインオーダーでたどった時訪れる節点の順に、 $D(4)$ の葉に左から右へと $T(3)$ の各節点を写像する。

エンコーディング $\hat{\varepsilon}_h$ $T(h) = (V, E)$ から $T_3(h+1) = (V, E')$ へのエンコーディング $\hat{\varepsilon}_h$ は次の様に定義される (図 3.3 参照)。

$\hat{\varepsilon}_h: x \mapsto x' \quad (x \in V, x' \in L, L \text{ は } T_3(h+1) \text{ の葉の集合とする。})$

ここで $l(x') = f_h(l(x))$ である。但し、

$$f_h(\alpha) = \begin{cases} 1 \cdot 2^h & (\alpha = \lambda) \\ 1 \cdot \beta \cdot 1^{|\gamma|} \cdot 2^{h-|\beta|-|\gamma|-1} \cdot \gamma \cdot 1^{1-|\gamma|} & (\alpha = 1 \cdot \beta \cdot \gamma, \beta \in \Sigma^*, \gamma \in \Sigma \text{ or } \beta = \gamma = \lambda) \\ 2 \cdot \beta \cdot 2^{|\gamma|} \cdot 1^{h-|\beta|-|\gamma|-1} \cdot \gamma \cdot 2^{1-|\gamma|} & (\alpha = 2 \cdot \beta \cdot \gamma, \beta \in \Sigma^*, \gamma \in \Sigma \text{ or } \beta = \gamma = \lambda) \end{cases}$$

である。

3.4 計算量の比較

主にホストグラフ H としたの DREE, TAP3 の計算量に関する

る比較を変換コスト C_1 および演算コスト C_2 を用いて行う。

(i) 変換コスト C_1

定理 3.1⁽¹⁾ $T(h)$ ($h > 0$) から $D(h+1)$ へのエンコーディング ε_h に対し, 任意の利用頻度 π に関し $2 \leq C_1(\varepsilon_h; \pi) \leq 3$ が成立する。

定理 3.2⁽⁴⁾ $T(h)$ ($h > 0$) から $TAP1 T_1(h+1)$ へのエンコーディング ε_h に対し, 任意の利用頻度 π に関し $C_1(\varepsilon_h; \pi) = 2$ が成立する。

定理 3.3 $T(h)$ ($h > 0$) から $T_3(h+1)$ へのエンコーディング $\hat{\varepsilon}_h$ に対し, 任意の利用頻度 π に関し $2 \leq C_1(\hat{\varepsilon}_h; \pi) \leq 3$ が成立する。

数学的帰納法を用いれば証明できるがここでは省く。

(ii) 演算コスト C_2

ゲストグラフ G を高さ h の完全二分木 $T(h)$ とした時の演算コスト C_2 に関する比較結果を表 3.1 に示す。ここで N は完全二分木 $T(h)$ 内の節点の総数とする。

謝辞

有益な御討論を頂いた大阪大学谷口健一助教授, 電気通信大学西沢輝泰助教授に深謝いたします。

文献

- (1) Rosenberg, A.L., Wood, D. and Galil, Z.: "Storage representations for tree-like data structures", Math. Syst. Th. 13, pp. 105-130 (1979).
- (2) Rosenberg, A.L.: "Data encodings and their costs", Acta

Inform. 9, pp. 273-292 (1978).

(3) Lipton, R.J., Eisenstat, S.C. and DeMillo, R.A.: "Space and time hierachies for classes of control structures and data structures", JACM 23, 4, pp. 720-732 (1976).

(4) 菊野, 杉原: "木構造データのエンコーディングについて", 電気四学会中支連大 (Dec. 1979).

H \ i		1: フレオーダ				2: イニオーダ				3: ポストオーダ				4: レベルオーダ			
T(h)	a	N				2N				2N				N			
	t	N				N				N				N			
	s	2N				4N				4N				2N			
D(h+1)	a	2N				2N				2N				2N			
	t	N				N				N				N			
	s	2N				4N				4N				2N			
T ₃ (h+1)	a	$N + \lfloor \frac{N}{2} \rfloor + \log(N+1)$				$N + \lfloor \frac{N}{2} \rfloor + \log(N+1)$				$N + \lfloor \frac{N}{2} \rfloor + \log(N+1)$				$N + \lfloor \frac{N}{2} \rfloor + \log(N+1)$			
	t	N				$N + \lfloor \frac{N}{2} \rfloor$				$N + \lfloor \frac{N}{2} \rfloor$				$\lfloor \frac{N}{2} \rfloor$			
	s	2N				3N-1				3N-1				N-2			

表 3.1 C_2 に関する比較

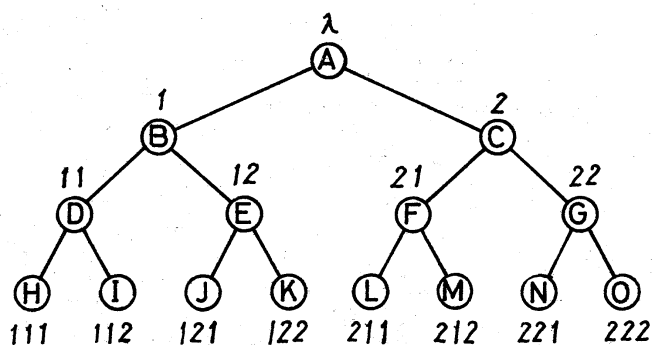


図 3.1 高さ 3 の完全 2 分木 $T(3)$

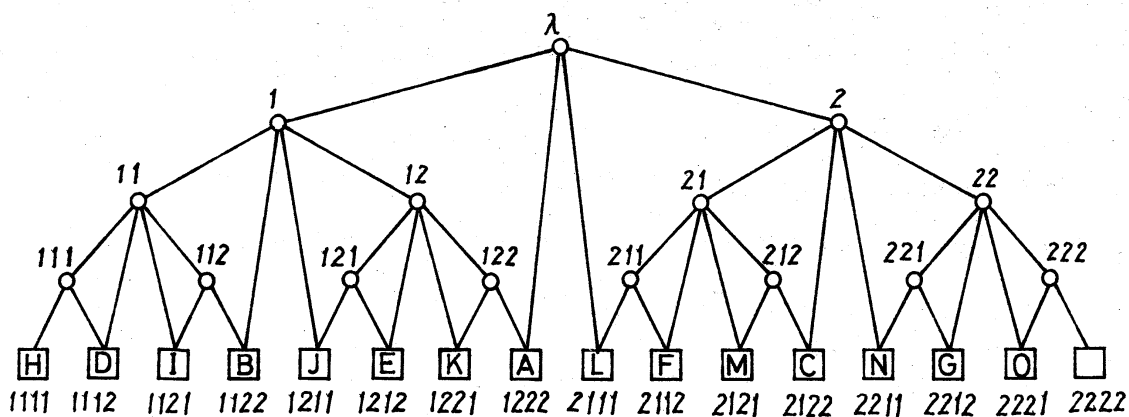


図 3.2 高さ 4 の DREE $D(4)$ とエンコーディング ε_3

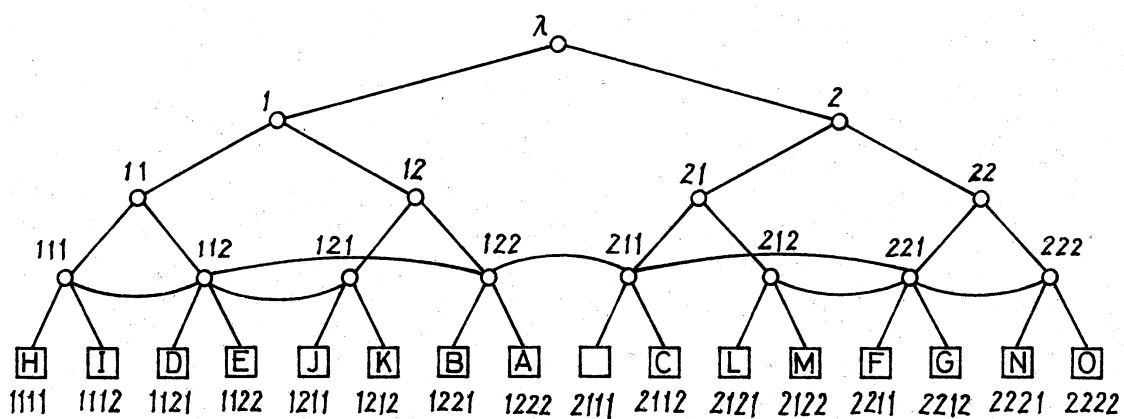


図 3.3 高さ 4 の TAP3 $T_3(4)$ とエンコーディング $\hat{\varepsilon}_3$