

〔論 文〕

## 生存時間解析における樹木構造接近法

衛 藤 俊 寿\*

\*日本文理大学工学部

### The Tree-Structured Approaches for Survival Data Analysis

Toshihisa ETO\*

\*School of Engineering, Nippon Bunri University

#### Abstract

The tree-structured approach has been positively used in survival studies in recent years because it hardly requires statistical assumptions and restrictions and can express the result visually.

In survival studies, from the viewpoint of effective and efficient treatment of a patient, searching for and evaluating the prognosis factors, which identify a patient's death risk, especially in the actual survival data analysis as typified by cancer research, is essential.

In this paper, we discuss the tree-structured approach, particularly the two-way split method, for survival data analysis.

First, we explain the concept of the tree-structured approach and summarize its development background. To clarify the features and applicable scope of the tree methods, we typify the tree-structured approach using six features and encapsulate typical methods corresponding to the types.

We also present an outline of the two-way split method in survival data analysis and an example of its application to gastric cancer patient data.

Finally, we conclude by providing some productive findings and discussing the problems in the two-way split method.

キーワード：生存時間解析，樹木構造接近法，2分岐法，胃癌データ解析，多分岐法

**Keywords** : survival data analysis, tree-structured approaches, two-way split method, gastric cancer data analysis, multi-way splits method

#### 1. はじめに

樹木構造接近法は，目的変数と説明変数の対で与えら

れる回帰型あるいは分類型のデータに適用され，その結果は説明変数のとりうる値の2個以上の分岐からなる階層構造の樹木図で表示される。樹木構造接近法は，Morgan & Sonquist (1963)<sup>[1]</sup>のAID (Automatic Inter-

action Detection: 自動交互作用検出法)に始まる。その名前が示すように、説明変数から目的変数を予測するための線形モデルに主効果のみを含めるか否か、すなわちそのモデルで説明変数の交互作用を考慮に入れるか否かを解明することがAIDの目的であった。その後、CHAID (Kass, 1980<sup>[21]</sup>), CART (Breiman *et al.*, 1984<sup>[31]</sup>), FACT (Loh & Vanichsetakul, 1988<sup>[41]</sup>), C4.5 (Quinlan, 1993<sup>[51]</sup>), QUEST (Loh & Shih, 1997<sup>[61]</sup>), FIRM (Hawkins, 1997<sup>[71]</sup>), CRUISE (Kim & Loh, 2001<sup>[81]</sup>)といった樹木構造接近法が開発される過程で、通常の回帰解析やクラスター解析に代わる方法論としてより広範囲の場面で利用されるようになった。

樹木構造接近法は、当該のデータに直面したときに何らかのモデルを想定せずにデータから直截に目的変数と説明変数の関係を評価できること、結果を説明変数の分岐からなる階層構造の樹木図で表現できるために説明変数からなる非線形構造あるいは交互作用構造を視覚的に解釈できること、説明変数の相対重要性を幹から枝にかける順序によって評価できること、説明変数の分岐で決定される説明変数空間に基づいて全個体を分類の「像」(プロフィール)を提示できること、アルゴリズムが簡明であることなどの利点をもっている。他方、最適分岐点を探索する手順をすべての説明変数について反復するために膨大な計算時間を要するといった欠点ももっている。

Ciampi *et al.* (1987)<sup>[9]</sup>やSegal (1988)<sup>[10]</sup>は、この樹木構造接近法を生存時間解析の場へ適用することを試みている。そこでは、生存時間解析で特有の観測値の中途打ち切りを考慮に入れて、検定統計量を最適な分岐を探索するための基準として採用している。

生存時間研究は、「生存時間分布の同定」、「生存時間分布の比較」、「予後因子の探索と予後の予測」のいずれかを目標とすることが多い(後藤他, 1988<sup>[11]</sup>)。とくに、癌研究に代表される生存時間研究の実地の場では、死亡のリスクに関与する予後因子を探索・評価することは、患者に効果的な治療を効率よく施すという見地から非常に重要である。従来、このような予後因子の探索・評価には、比例ハザード・モデル(Cox, 1972<sup>[12]</sup>; 後藤・松原, 1982<sup>[13]</sup>; Goto & Matsubara, 1996<sup>[14]</sup>)が汎用されてきた。比例ハザード・モデルでは、共変量の効果が対数ハザードに対して加法項で表され、個体の(解析対象集団に対する)相対ハザードの大きさは共変量の大きさに比例することが仮定されている。樹木構造接近法は、このような仮定や制約を考慮に入れることがなく、しかも結果を視覚的に表現できることから、近年、生存時間

解析の場で積極的に用いられている(例えば、松原他(1990)<sup>[15]</sup>を参照)。

本稿では、生存時間データ解析のための樹木構造接近法、特に2分岐法について論じる。ここでは、樹木構造接近法の概略とその生存時間データ解析への応用について検討する。まず、樹木構造接近法を6つの特徴から類型化し整理する。次に、2分岐法に分類される樹木構造接近法の生存時間解析への応用について、その方法を述べ、2分岐法を実際のデータセットへ適用した事例を紹介する。さらに、2分岐法の問題点を提示し今後の展望を考察する。

## 2. 樹木構造接近法とその生存時間解析への応用

### 2-1 樹木構造接近法の類型

樹木構造接近法は、2分岐樹木と多分岐樹木のグループに大別できる。AID, CART, QUEST, CRUISEは前者のグループに属する。後者に属するのは、FACT, C4.5, CHAID, FIRMである。FACTは、各分岐においてクラスごとに一つのふしを生成する。C4.5は、選択された変数が数値ならば2分割を生成し、カテゴリーならばふしはC個の子ふしに分割する(Cはカテゴリー数である)。CHAIDは、C4.5に類似しているが、いくつかのふしを結合する追加過程をもつ(この過程は「説明変数のグループ化」と呼ばれる)。FIRMははじめにそれぞれの範囲を10区間に分けることによって、CHAIDの概念を数値変数に拡張した。

Morgan & Sonquist (1963)<sup>[11]</sup>の最初のAIDは、最小二乗回帰によりステップワイズ最適樹木構造を構成した。ただし、AIDでは、説明変数がカテゴリカル変数に限定されているだけでなく、停止基準が解析者に委ねられているため、多分に恣意的である。この平方和を基盤とした最適樹木構造の探索方式は、CARTの最小二乗偏差あるいは最小絶対偏差という分岐基準に引き継がれる。AIDはその後、その手順において分割または併合の必要性を判断するために、2グループを比較するための適当な検定統計量を必要とした。

2グループ用の統計量としてt統計量を採用したAIDの実行方式はXAID (Heymann, 1981<sup>[16]</sup>)である。従属変数が間隔尺度で与えられていても正規性を仮定することが不適当である場合には、対応するノンパラメトリック統計量に基づく方式が利用できる。例えば、多分岐に対するKruskal-Wallis統計量と、その2グループへの特殊化手法であるMann-Whitney統計量が利用できる(Worsley, 1977<sup>[17]</sup>)。目的変数が名義尺度の場合、

表 1. 樹木構造接近法の類型

T1	T2	T3	T4	T5		樹木構造 接近法	T6
				分割変数選択	分割点選択		
連続	連続	2分岐	同期	最小二乗偏差 (LSD)		CART	—
		多分岐	同期	カイ二乗検定		FIRM	10区間の組合せ
	カテゴリー	2分岐	同期	平方和		AID	カテゴリー組合せ
		多分岐	同期	ANOVA F 検定		CHAID	カテゴリー組合せ
カテゴリー	連続	2分岐	同期	ANOVA F 検定	LDA	FIRM	カテゴリー数
			非同期	最小二乗法 (LSD)		CART	—
	多分岐	非同期	ANOVA F 検定	QDA	QUEST	—	
			ANOVA F 検定 (+ブートストラップ)	LDA (ベキ変換)	CRUISE	—	
	カテゴリー	2分岐	非同期	カイ二乗検定	QDA	QUEST	—
				多分岐	同期	カイ二乗検定	
		非同期	カイ二乗検定	LDA	C4.5	カテゴリー数	
			カイ二乗検定 (+ブートストラップ)	LDA (ベキ変換)	CRUISE	—	

【T1】 目的変数の属性(連続またはカテゴリー), 【T2】 説明変数の属性(連続またはカテゴリー), 【T3】 親ふしから生成される子ふしの数, すなわち, 分岐数(2分岐または多分岐), 【T4】 分岐変数の選択と分岐点の選択を同時に行うか否か(同期または非同期), 【T5】 分岐判定のための統計量, 【T6】 説明変数のグループ化

THAID (Morgan & Messenger, 1973<sup>[18]</sup>) は最頻クラスへの最適予測を意図したときに, 正しくクラス分けされる標本の比率という基準を採用した。一方, CHAID は分割表に対するカイ二乗検定統計量とその基盤となっている。

本稿では, このような樹木構造接近法の開発経緯に基づきその特徴を以下の特性によって分類・整理した。

- T1. 目的変数の属性 (連続またはカテゴリー)
- T2. 説明変数の属性 (連続またはカテゴリー)
- T3. 親ふしから生成される子ふしの数, すなわち  
分割数 (2分岐または多分岐)
- T4. 分割変数の選択と分割点の選択を同時に行うか  
否か (同期または非同期)
- T5. 分割判定のための統計量
- T6. 説明変数のグループ化

分類結果を表1に示す。とくに, 樹木構造接近法を分岐過程から眺めると, どの説明変数を順に用いて分割するかの問題 (説明変数の選択) と, その分割点 (データ点) が最適に選べるかの問題 (分割点の選択) がある。本稿では, 説明変数の選択と分割点の選択を同時に行う方法を同期型, 別々に行う方法を非同期型とする。

### 2-2 生存時間解析における2分岐法

非中途打ち切り観測値を含む生存時間と共変量の関係を樹木構造で表現する樹木構造表現法は, 一般には「回帰樹木」と呼称されている。回帰樹木は, 通常の回帰解

析の方法と同じく, 応答変数と説明変数の対で与えられるデータに適用されるが, この方法では回帰解析のモデルの線形性や誤差の仮定を満たさずともよいという利点を有している。いま, 連続な応答  $y_i$  と, それに対応する  $p$  個の説明変数  $x_{i1} \dots x_{ip}$  が  $n$  組 ( $n$  個の個体について) 観測されているとする。このとき, 回帰樹木は次の4個の要素から構成され, その結果は樹木図で表示される。

- (1) 説明変数  $x_j$  とその説明変数のとり得る値  $c_j$  に対して「 $x_j \leq c_j$  であるか否か」による標本の2分割 ( $j = 1, \dots, p$ ): この分割で形成される二つの部分標本を「ふし」と呼ぶ。
- (2) ふし  $S$  の2分割に対する評価基準  $\varphi(S)$ : ふし  $S$  の2分割によって  $S_L$  と  $S_R$  の二つのふしが形成されるとする。この分割の効果の評価基準として, 通常, 応答  $y$  のふし内偏差平方和の減少量, すなわち  $\varphi(S) = RSS(S) - RSS(S_L) - RSS(S_R)$  が用いられる。ここに  $RSS(S)$  は応答  $y$  のふし  $S$  内の偏差平方和, および  $RSS(S_L)$  と  $RSS(S_R)$  はそれぞれ, ある説明変数のある区分点による左枝 ( $L$ ) と右枝 ( $R$ ) への分割で生じる2個のふし, すなわち部分標本  $S_L$  と  $S_R$  のふし内偏差平方和を表す。実際には, すべての説明変数とそれらの説明変数のとり得るすべての区分点について  $\varphi(S)$  を計算し, その中で最大の  $\varphi(S)$  を与える説明変数とその区分点がふし  $S$  での最良の分割として選択される。
- (3) 分割の停止基準: 対象となるふしでの標本サイズが予め規定したサイズよりも小さくなるか, または

$\varphi(S)$ が予め規定した数値よりも小さくなったときに停止する。この基準によって、樹木の大きさが決まる。分割が停止したときのふしを「終結ふし」と呼ぶ。

- (4) 各ふし（部分標本）における応答の要約統計量：この要約統計量は最終的に得られた回帰樹木を解釈するのに用いられる。

なお、回帰樹木法についての詳細は、Breiman *et al.* (1984)<sup>[3]</sup>を参照されたい。

Segal (1988)<sup>[10]</sup>は、この回帰樹木の中途打ち切り観測値を含む生存時間データへの応用を試みている。このとき、上記のふし内平方和による分割評価基準の代りに、生存時間分布の位置のふし間の差を測度とする統計量を導入している。ここでは、この測度として Tarone & Ware (1977)<sup>[19]</sup>で与えられた2標本順位検定統計量を利用する。

いま、2個の標本があり、それらの併合標本での生存時間を昇べき順に  $t(1) < t(2) < \dots < t(k)$ 、第  $i$  時点でのリスク集合のサイズを  $n_i$ 、死亡数を  $m_i$  とする ( $i = 1, \dots, k$ )。また、一方の標本のリスク集合のサイズを  $n_{1i}$ 、死亡を  $d_{1i}$  で表す。同様に、他方の標本に対して  $n_{2i}$  と  $d_{2i}$  を定義する。したがって、 $n_{1i} + n_{2i} = n_i$ 、 $d_{1i} + d_{2i} = m_i$  である。このとき、2個の標本の生存時間分布を比較するための検定統計量は

$$u = \sum_{i=1}^k w_i \left( d_{1i} - \frac{m_i n_{1i}}{n_i} \right) \left\{ \sum_{i=1}^k w_i^2 \frac{m_i (n_i - m_i) n_{1i} n_{2i}}{n_i^2 (n_i - 1)} \right\}^{\frac{1}{2}}$$

で与えられる (Tarone & Ware, 1977<sup>[19]</sup>)。  $u$  は2標本の生存時間分布が等しいとする帰無仮説のもとで近似的に平均0と分散1の正規分布に従う。ここに、 $w_i$  は重みを表し、統計量  $u$  は  $w_i = 1$  のときログランク検定 (Peto & Peto, 1972<sup>[20]</sup>)、 $w_i = n_i$  のとき Gehan (1965)<sup>[21]</sup> の一般化 Wilcoxon 検定、 $w_i = \sqrt{n_i}$  のとき Tarone & Ware (1977)<sup>[19]</sup> の検定、 $w_i = S_i$  のとき Peto & Peto (1972)<sup>[20]</sup> と Prentice (1978)<sup>[22]</sup> の一般化 Wilcoxon 検定の各統計量に相当する。ただし、

$$S_i = \prod_{j=1}^i n_j / (n_j + 1)$$

は第  $i$  番目の死亡時間での Kaplan-Meier 推定値である。この検定統計量を分割評価基準とする回帰樹木は次のように構成することができる。

- (1) 任意のふし  $S$  において、説明変数  $x_j$  について可能なあらゆる2分割を考える。
- (2) その可能な2分割で得られる最大の検定統計量を  $u_j$ 、その  $u_j$  を与える区分点を  $c_j$  とする。
- (3) 各  $j (= 1, \dots, p)$  に対して、最大の  $u_j$  の添字をもつ説明変数をふし  $S$  での分割説明変数、および対応する  $c_j$  をその説明変数における区分点とする。
- (4)  $u_j$  が予め規定した有意水準で有意でなくなるか、ふしの大きさ（標本サイズ）が予め規定したサイズよりも小さくなったときに分割を停止する。

上記の手順で構成される回帰樹木において、アルゴリズムの性質から同一ふしで分割された終結ふし間の生存時間分布は異なると考えられる。しかし、終結ふしのあらゆる部分標本が異なる生存時間分布をもつ保証は何もない。したがって、最終的に異なる生存時間分布をもつ部分標本への分割が望まれるときには、終結ふしを再併合することが必要となる。この併合は分割の逆過程をたどることで行える。すなわち、終結ふしのあらゆる組み合わせ対について検定統計量を求める。それらの統計量について予め規定した有意水準で有意でない最小の統計量をもつ組み合わせ対を併合する。併合された対を一つのクラスターと考え、このクラスターと終結ふし間の検定統計量を求める。これらの統計量と先に併合されなかった組み合わせ対間の統計量の中で、有意でない最小の統計量をもつクラスターと終結ふし、あるいは二つの終結ふしを再度併合する。この手順を最小の検定統計量が有意になるまで繰り返す。

### 3. 適用事例

愛知県がんセンター、癌研究会附属病院、国立がんセンター、大阪府立成人病センターの4施設で胃癌の術後補助療法に関する第2次共同研究によって胃癌症例データが収集された (Yamamura *et al.*, 2002<sup>[23]</sup>)。データは、ESAC 第2次研究の投薬計画における A 群（外科手術群）および B 群（薬剤投与群）について観測された555症例である。これらの症例について、術後の生死および生存月数が観測されていた。

さらに、各症例について、影響因子の候補として、

- (1) 性、(2) 年齢、(3) 占拠部位、(4) 腫瘍径、(5) 漿膜面浸潤 (S)、(6) リンパ節転移 (N)、(7) 肉眼型、(8) 郭清程度 (R)、(9) 薬剤強度が観測されていた。

これらの因子候補のカテゴリーを表2に示す。なお、

表2. 観測された項目のカテゴリー区分

予後因子候補	カテゴリー	予後因子候補	カテゴリー
(1) 性	男 (692 例), 女 (381 例)	(6) リンパ節転移	N0 (278 例), N1 (484 例), N2 (218 例), N3 (32 例), N4 (61 例)
(2) 年齢	~29 歳 (5 例) 30~49 歳 (222 例) 50~69 歳 (717 例) 70~ (129 例)	(7) 肉眼型	O (107 例), I (48 例), II (260 例), III (320 例), IV (140 例), V (198 例)
(3) 占拠部位	A (388 例), M (358 例), C (232 例), AMC (95 例)	(8) 郭清程度	R <sub>0</sub> (11 例), R <sub>1</sub> (42 例), R <sub>2</sub> (773 例), R <sub>3</sub> (247 例)
(4) 腫瘍径	4cm 未満 (297 例) 4cm 以上 8cm 未満 (507 例) 8cm 以上 (269 例)	(9) 腹膜播種	P <sub>0</sub> (997 例), P <sub>1</sub> (25 例), P <sub>2</sub> (31 例), P <sub>3</sub> (20 例)
(5) 漿膜面浸潤	S0 (429 例), S1 (133 例), S2 (425 例), S3 (86 例)	(10) 薬剤強度	1 未満 (959 例) 1 以上 2 未満 (111 例) 2 以上 (3 例)

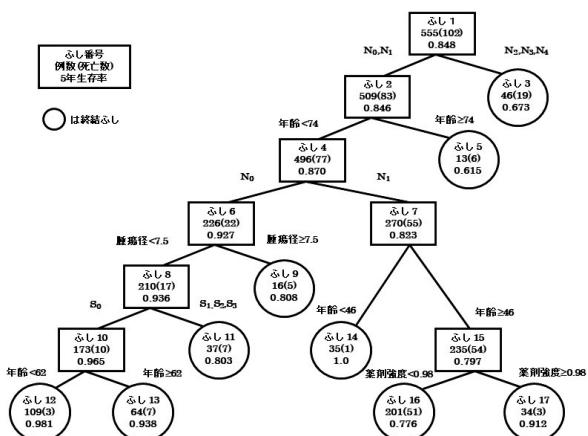


図1. 樹木図

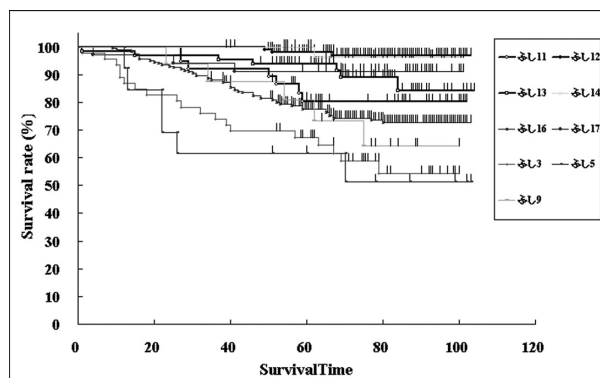


図2. 終結ふしに対応する Kaplan-Meier プロット

解析では、年齢、腫瘍径、薬剤強度は連続量として取り扱った。

本研究の症例群について、胃癌手術症例の予後因子を探索するために、術後生存時間と因子候補の関係を Segal (1988)<sup>[10]</sup>の生存時間 CART 法で樹木表現した。

樹木の構成では、各ふしの分割基準にはログランク統計量を用い、分割で生じた部分集合内の症例数が10例以下になったとき、またはログランク統計量の p 値が 0.05 を超えたときに、ふしの分割を停止させた。

その結果得られた樹木図を図1に示す。図中の各「ふし」の上段、中段、下段のそれぞれに、ふし番号、症例数(死亡数)、5年生存率を示す。このときの終結ふしに対応する症例群における Kaplan-Meier プロットを図2に示す。

この結果から、症例群の術後予後には、リンパ節転移 (N)、年齢、腫瘍径、漿膜面浸潤 (S)、薬剤強度がこ

の順に術後生存率への影響度が高いことが示唆される。終結ふしにおける5年生存率から、ふし番号14 (リンパ節転移の程度が N<sub>1</sub> の46歳未満) の患者群 (35例, 5年生存率: 100%) と、ふし番号12 (リンパ節転移の程度が N<sub>0</sub> の62歳未満の症例のうち腫瘍径が7.5cm 未満かつ漿膜面浸潤が S<sub>0</sub>) の患者群 (100例, 5年生存率: 98.1%) の術後生存時間が長い傾向にあった。逆に、ふし番号3 (リンパ節転移の程度が N<sub>2</sub>, N<sub>3</sub>, N<sub>4</sub>) の患者群 (46例, 5年生存率: 67.2%) と、ふし番号5 (リンパ節転移の程度が N<sub>0</sub>, N<sub>1</sub> の74歳以上の患者群 (13例, 5年生存率: 61.5%) の予後の悪いことが示唆される。

また、薬剤効果を表すと考えられる薬剤強度が関与していた症例群はふし番号15 (235例) であり、薬剤強度が0.98以上と0.98未満の群の間に、5年生存率で約0.14の差が認められた。終結ふしに対応する患者 (クラスター) プロフィールを表3に示す。

表3. 終結ふしに対応する患者（クラスター）プロフィール

ふし番号 例数 (5 生率)	リンパ節転移				年齢			腫瘍径 7.5	漿膜面浸潤		薬剤強度 0.98
	N <sub>0</sub>	N <sub>1</sub>	N <sub>2</sub>	N <sub>3</sub>	46	62	74		S <sub>0</sub>	S <sub>1</sub>	
ふし14 35例 (1.0)		○			←						
ふし12 109例 (0.981)	○				←	←		←	○		
ふし13 64例 (0.938)	○					→		←	○		
ふし17 34例 (0.912)		○			←	→					→
ふし9 16例 (0.808)					←			→			
ふし11 37例 (0.803)	○							←	○		
ふし16 201例 (0.776)		○			←	→					←
ふし3 46例 (0.673)			○	○							
ふし5 13例 (0.615)	○	○						→			

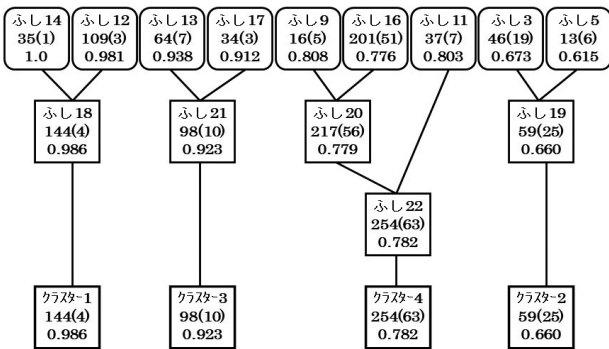


図3. 終結ふしの併合過程における樹木図

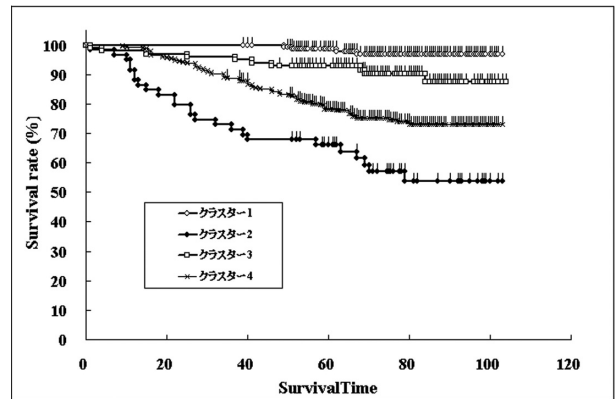


図4. 併合後の各クラスターのKaplan-Meierプロット

さらに、分割の逆過程をたどることにより終結ふしを併合した（併合ふし番号が小さい順により早く併合したことを表す）。ここでは、各ふしの併合基準にはログランク統計量を用い、ログランク統計量のp値が0.05より小さくなったときにふしの併合を停止させた。

終結ふしの併合過程における樹木図を図3に、併合後の各クラスターに対応するKaplan-Meierプロットを図4に示す。これらの結果から、最終的にクラスターは4つに分類され、これらのクラスターは死亡に対するリスクの大きさが相当に異なることが示唆される。

4. 結びに代えて

本稿では、生存時間研究における予後因子の探索を樹

木構造接近法によって探索した。まず、樹木構造接近法の適用場面を整理し、その類型を示した。次に、樹木構造接近法の生存時間研究への応用方式として2分法を概略した。さらに、2分岐法を実際のデータセットへの適用事例を紹介した。

樹木構造接近法における分岐アルゴリズムは2つのグループ、すなわち、2分岐法と多分岐法に大別されるが、非中途打ち切り観測値を対象とする樹木法では、CART, AID, QUESTは前者に属し、FACT, C4.5, CHAID, FIRM, CRUISEは後者に属する。一方、中途打ち切り観測値を対象とする樹木法については、現在、いくつかの2分岐法が提案されているが、多分岐法は開発されていない。

多分岐をもつ樹木法は常に2分岐樹木として再描画す

ることができる。そのため、多分岐法の利点はないようにみえるが、この結論は真ではない。結果が階層構造で表現される樹木法では、分岐の順序や分岐点が僅かに変わることによって樹木自体が大きく変化する恐れがある。このとき、同一の条件と状況で得られたデータのもとでは、僅かな影響要因に頑健な、安定した樹木を構成することが望まれる。AIDのような二分岐法では、その単純化指向によって計算負荷が軽くなり、その手順に関連した統計的推測も単純化されるが、一方でこの単純化は最終的な樹木を作るのに多段階を必要とするという欠点を持ち、不体裁で誤解を招きやすい樹木図の得られることがある (Hawkins, 1982<sup>[24]</sup>)。多段階の樹木を構成する二分岐法は、不安定な樹木を構成する危険性が高い。この観点から、不必要に多段階の樹木構造を構成しない多分岐法は、二分岐法の妥当性を診断する一つの方法となりうる。そうすることで、樹木構造接近法は、解釈の容易性や樹木構造の妥当性を通してより広範囲の研究者に受け入れられることになるはずである。とくに、生存時間研究の研究者は、二分岐より (二分岐を含む) 多分岐の概念を理解しやすく、実地の現象に沿った行動をとりやすいと予想される。今後は、生存時間データ解析における多分岐型樹木構造接近法を開発し、実地の生存時間データへの適用や性能を評価することで、その有用性を検証することが必要であろう。

#### 参考文献

- [1] Morgan, J.A. and Sonquist, J.N. (1963). Problems in the analysis of survey data and a proposal. *J. Amer. Statist. Assoc.*, 58, 415-434.
- [2] Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Appl. Statist.*, 29, 119-127.
- [3] Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C.J. (1984). Regression trees. *Classification and Regression Trees*. Ch. 8, 216-265, Wadworth.
- [4] Loh, W.Y. and Vanichsetakul, N. (1988). Tree-structured classification via generalized discriminant analysis (with comments). *J. Amer. Statist. Assoc.*, 83, 715-728.
- [5] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.
- [6] Loh, W.Y. and Shih, Y.S. (1997). Split selection methods for classification trees. *Statistica Seneca*, 7, 815-840.
- [7] Hawkins, D.M. (1997). FIRM : Formal inference-based recursive modeling, PC version, Release 2.1. *Technical Report*, 546, School of Statistics, University of Minnesota.
- [8] Kim, H and Loh, W.Y. (2001). Classification trees with unbiased multiway splits. *J. Amer. Statist. Assoc.*, 96, 589-604.
- [9] Ciampi, A., Chang, C.H., Hoggs, S. and McKinney, S. (1987). Recursive partition : A versatile method for exploratory data analysis in biostatistics. ed. by Umphrey, G., *Proceedings from Joshi Festschrift*, North Holland, Amsterdam.
- [10] Segal, M.R. (1988). Regression trees for censored data. *Biometrics*, 44, 35-47.
- [11] 後藤昌司・惣田隆生・佐久間昭 (1988). 生存時間研究での統計学的諸問題. *癌生時研誌*, 8, 75-80.
- [12] Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc.*, B 34, 187-220.
- [13] 後藤昌司・松原義弘 (1982). 比例ハザードモデルとその周辺. *応用統計学*, 11 (1), 1-26.
- [14] Goto, M. and Matsubara, Y. (1996). The proportional hazard model and its marginals. *The Best on Quality* [Vol. 7], ed. by John D. Hromi, ASQC *Quality Press*, 127-164.
- [15] 松原義弘・寺西孝司・後藤昌司 (1990). 癌の予後評価における樹木構造表現. *癌生時研誌*, 10, 15-20.
- [16] Heymann, C. (1981). XAID—an extended automatic interaction detector. *Internal Report*, CSIR.
- [17] Worsley, K. J. (1977). A non-parametric extension of a cluster analysis method by Scott and Knott. *Biometrics*, 33, 532-535.
- [18] Morgan, J.A. and Messenger, R.C. (1973). THAID—a sequential analysis program for the analysis of nominal scale dependent variables. SRC, *Inst. Soc. Res.*, University of Michigan.
- [19] Tarone, R.E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, 64, 156-160.
- [20] Peto, R. and Peto, J. (1972). Asymptotically efficient rank-invariant test procedures. *J. Roy. Sta-*

- tist. Soc.*, A (135), 185–198.
- [21] Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52, 203–223.
- [22] Prentice, R. L. (1978). Linear rank tests with right-censored data. *Biometrika*, 65, 167–179.
- [23] Yamamura, Y., Nakajima, T., Ohta, K., Nashimoto, A., Arai, K., Hiratsuka, M., Sasako, M., Kodera, Y. and Goto, M. (2002). Determining prognostic factors for gastric cancer using the regression tree method. *Gastric Cancer*, 5 (5), 201–207.
- [24] Hawkins, D. M. (1982). Topics in Applied Multivariate Analysis. *Cambridge University Press*, London.

---

(2022年12月20日受理)