

# 共通構造の数え上げによる半構造データカーネルの設計

久保山哲二<sup>†</sup> 申吉浩<sup>††</sup> 鹿島久嗣<sup>†††</sup> 平田耕一<sup>††††</sup>

<sup>†</sup> 東京大学 国際・産学共同研究センター 〒153-8505 東京都目黒区駒場 4-6-1

<sup>††</sup> 東京大学 先端科学技術研究センター 〒153-8904 東京都目黒区駒場 4-6-1

<sup>†††</sup> 日本 IBM 東京基礎研究所 〒242-8502 大和市下鶴間 1623-14

<sup>††††</sup> 九州工業大学 情報工学部 〒820-8502 福岡県飯塚市川津 680-4

E-mail: <sup>†</sup>kuboyama@ccr.u-tokyo.ac.jp, <sup>††</sup>kshin@mpeg.rcast.u-tokyo.ac.jp

<sup>†††</sup>hkashima@jp.ibm.com, <sup>††††</sup>hirata@ai.kyutech.ac.jp

**あらまし** 機械学習の手法としてカーネル法が広く用いられるようになり、半構造データや構文木、RNA の二次構造などの木構造を対象としたカーネル関数が数多く提案されている。鹿島らは、畳み込みカーネルと呼ばれる木カーネルの設計において、木の部分構造を取り出す際の規則をどこまで緩めることができるかという問題に取り組み既存の木カーネルの拡張を行った。本研究では、鹿島らによる木カーネルが、木写像と呼ばれる木の共通部分パターンを表す構造の数え上げになっていることを示し、木カーネルを統一的に記述・設計するための一般的な枠組みを示す。この枠組みのもとで、鹿島らによる既存の木カーネルを拡張し、より柔軟に部分構造を数え上げることができる木カーネルを提案する。また、木写像の数え上げ関数が理論的にも正しいカーネル関数となっていることを実際に特徴空間を示すことにより証明し、提案したカーネル関数の性質を明らかにする。

**キーワード** 畳み込みカーネル、木カーネル、数え上げ問題、木写像、木の編集距離

## Kernel Design for Semistructured Data based on Counting Common Subtree Patterns

Tetsuji KUBOYAMA<sup>†</sup>, Kilho SHIN<sup>††</sup>, Hisashi KASHIMA<sup>†††</sup>, and Koichi HIRATA<sup>††††</sup>

<sup>†</sup> Center for Collaborative Research, The University of Tokyo,  
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

<sup>††</sup> Research Center for Advanced Science and Technology, The University of Tokyo,  
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904 Japan

<sup>†††</sup> Tokyo Research Laboratory, IBM Japan, Ltd.,  
1623-14 Shimotsuruma, Yamato-shi, Kanagawa, 242-8502 Japan

<sup>††††</sup> Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology  
680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

E-mail: <sup>†</sup>kuboyama@ccr.u-tokyo.ac.jp, <sup>††</sup>kshin@mpeg.rcast.u-tokyo.ac.jp

<sup>†††</sup>hkashima@jp.ibm.com, <sup>††††</sup>hirata@ai.kyutech.ac.jp

**Abstract** A number of kernel functions for tree structures have been proposed for machine learning on semistructured data, parsing trees, RNA secondary structures, and so on. Kashima et al. have addressed a problem of relaxing a restriction in extracting substructures from trees for designing a more general tree kernels than previously proposed ones. In this paper, we extend the tree kernels due to Kashima et al., and propose a general framework for designing tree kernels based on convolution kernels. This framework formulates the designing problem of tree kernels as the counting problem of tree mappings between two trees. By using our framework, we propose several new tree kernels according to the class hierarchy of tree mappings. Moreover, we reveal the feature spaces of these kernels.

**Key words** Convolution Kernels, Tree Kernels, Counting Problems, Tree Mappings, Tree Edit Distance

## 1. はじめに

インターネット上の HTML 文書や、XML に基づくデータの増大によって、これらのいわゆる半構造データから、必要な情報を効率よく獲得するための手法の重要性が増している。半構造データは、木構造データとして捉えることができ、自然言語解析における構文解析木、バイオインフォマティクスにおける進化系統樹、RNA の二次構造、糖鎖の分岐構造などとともに、木構造を対象としたアルゴリズムを用いた比較や機械学習の手法が広く研究されている。

一方で、機械学習の手法として、1990 年代後半からカーネル法が脚光を浴びるようになった。カーネル法は、学習タスクを受け持つカーネルマシンと、入力データをカーネルマシンが扱うことのできる特徴空間へ埋め込むためのカーネル関数から構成される。代表的なカーネルマシンであるサポートベクターマシンは、次元に依存しない汎化性能が理論的、実験的に示されている。

カーネル関数は、2 つの特徴ベクトルの内積であり、内積さえ何らかの方法で計算することができれば、明示的に対象データの特徴ベクトルを構成することなく高次元の特徴ベクトルを扱うことが可能になる。また、2 つのデータの類似度を計算する適当な関数が、ある条件 (半正定値性) を満たすときには、その関数をカーネル関数として用いることができる。しかし、この条件は、どのようなデータ集合であっても線形空間に埋め込めることを保証しているにすぎず、学習器としてのカーネル法の性能を保証するものではない。

カーネル関数の設計については、これまでに様々な方法が提案されてきた。木構造やグラフ構造をはじめとする離散的な構造をカーネル法によって扱うための一般的な枠組みとしては、**畳み込みカーネル (convolution kernel)** [1] が最もよく知られている。構造を持ったデータの特徴は、その部分構造が担っていると考えるに基づき、構造データ間の類似性を、その部分構造間の類似性として再帰的に定義する方法である。

サポートベクターマシンに代表されるカーネルマシンの優れた性能と、一般的なカーネル関数設計の枠組みによって、様々なデータを対象とした学習器を比較的容易に設計できる道が開けた。このため、配列カーネル、木カーネル、グラフカーネルなど、離散構造に対しても様々なカーネル関数が提案されている [2]。このようなカーネル関数の設計ブームも、基本的な構造に対するカーネル関数の提案が出尽くし、ようやく落ち着きを見せてきた観がある。これまでに提案されてきた畳み込みカーネルに基づくカーネル関数の中には、背後にある特徴空間の意味をよく考えずに設計されてきたものも散見される。設計したカーネル関数の実効性は、つまるところその対象とする問題への性能評価によって示すしかないが、よりよいカーネル関数設計のためには、その背後にある意味を正しく理解することが不可欠である。

木の比較や照合のアルゴリズムにおいては、木構造間の距離を測るための最も一般的な方法として**木の編集距離**が知られている。木の編集距離は、**木写像 (tree mapping)** という構造の

最適化問題として定式化できる [3], [4]。木写像は 2 つの木構造間の照合パターンを表現するための構造であり、対象とする問題の性質に応じて様々なクラスが提案されている。また、木写像のクラスは木の編集距離の計算量と密接な関係があることがわかっており、木写像のクラスの些細な違いが、木の比較の意味や計算量に大きな影響を及ぼすため、効率を考える上でも重要な概念である。

本稿では、次の 3 つの結果を示す。

- 2 つの木の間に存在する木写像の数を、既知のさまざまなクラスの木写像に対して数え上げることができるような一般的なアルゴリズムを示す。

- 提案したさまざまな木写像のクラスに対する数え上げ関数が、カーネル関数としての条件を満たしていることを、実際の特徴空間を示すことにより証明する。

- このようにして得られた木カーネルが、鹿島らによる既存研究 [5], [6] の拡張になっていることを示す。

## 2. 木構造

本稿では、ノードにラベルがあり、かつ子ノード間に順序のある根つき木を対象とし、これを木と呼ぶ。木  $T$  のすべてのノードからなる集合を  $V(T)$  で表す。また、ノード  $x$  と、その祖先  $y$  との関係を、 $x \leq y$  と表記する (ノード  $x$  自身も  $x$  の祖先であることに注意)。とくに、 $x < y$  のとき、 $y$  を**真の祖先**という。ラベル空間を  $\Sigma$  とし、木の各ノードにラベルを割り当てる関数を  $\ell$  とする。木  $T$  の根ノードを  $\text{root}(T)$ 、ノード  $x$  の子ノードの集合を  $\text{ch}(x)$ 、木  $T$  に含まれるノード  $x$  について、 $x$  を祖先とするすべてのノードを含む部分木を  $T(x)$  とそれぞれ表記する。

木  $T$  に含まれるノードの数を、木の大きさと定義し、 $|T|$  で表す。

木  $T$  とそのノード集合  $V' \subseteq V(T)$  が与えられたとき、 $V'$  のすべての要素の祖先であるようなノードを  $V'$  の**共通祖先**という。 $V'$  の共通祖先の中で最も葉に近い位置にあるノードを、 $\text{lca}(V')$  と表記する。また、 $x \sim y$  により、 $\text{lca}(\{x, y\})$  を表す。

本稿では木の記述に次のような表現を用いる。木を根ノードと互いに素な木の順序列によって表記する。このような木の順序列を**森 (forest)** といい、次のように再帰的に定義する。

- (1)  $T$  が木であるとき、 $(T)$  は森である。
- (2)  $T$  が木、 $F$  が森であるとき、 $(T \circ F)$  は森である。

よって、木の列  $T_1, \dots, T_n$  から構成される森は、 $F = (T_1 \circ (T_2 \circ (T_3 \circ \dots \circ (T_n))))$  と表すことができる (以後、簡単のために括弧を省略する)。このとき、 $F = \prod_{i=1}^n T_i$  という表記も用いる。空の森は  $\emptyset$  で表す。また、森  $F$  を構成する各木の根ノードの直上の親となるように新たに根ノード  $v$  を加えることによってできる木を  $v(F)$  と表記する。2 つの森  $F_1$  と  $F_2$  の木の列を順序を保ったまま結合して得られる森を  $F_1 \bullet F_2$  と表記する。

## 3. 木の編集距離と木写像

2 つの木構造間の照合の構造をあらわすために、**木写像**とい

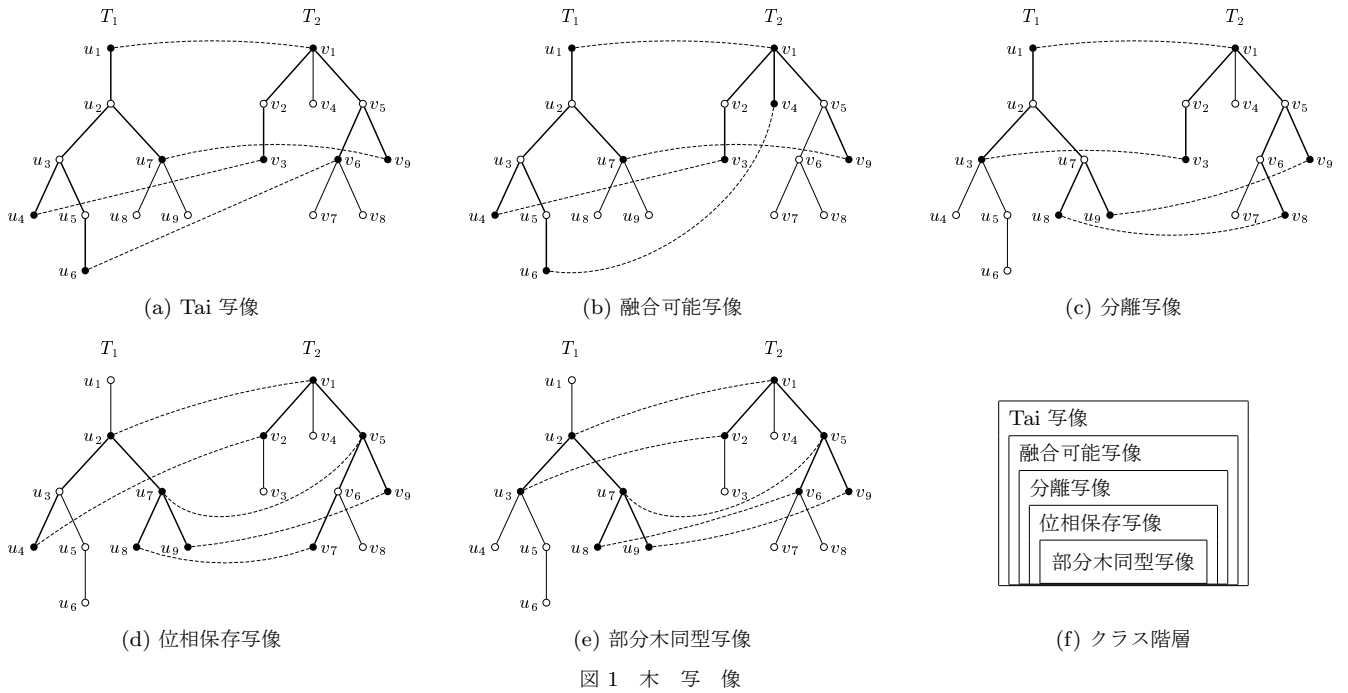


図 1 木写像

う概念を用いる。木写像とは、2つの木構造間のノード間の対応関係を表す集合である。

### 3.1 Tai 写像

まず、木写像の中でもっとも一般的な Tai 写像の定義を示す。Tai 写像は、2つの木の部分ノード集合を考えたとき、ノード間で先祖・祖先、左右の関係が保たれているようなノードの一对一対応を表す構造である。

**定義 1** (Tai 1979 [3]). 木  $T_1$  から  $T_2$  への **Tai 写像**  $M$  とは、次の条件を満たす  $V(T_1) \times V(T_2)$  の部分集合である。任意のノード対  $(x_1, x_2), (y_1, y_2) \in M$  について、

- (1)  $x_1 \leq y_1 \Leftrightarrow x_2 \leq y_2$ ,
- (2)  $x_1$  は  $y_1$  の左側  $\Leftrightarrow x_2$  は  $y_2$  の左側。

ここで、ノード  $x$  がノード  $y$  の左側であるとは、ノード  $z = x \smallfrown y$  が、 $x$  と  $y$  の真の祖先であり、 $z$  から  $x$  へのパス上にある  $z$  の子ノードが、 $y$  へのパス上にある  $z$  の子ノードの左側にあることを意味するものとする。

### 3.2 木写像の最適化問題としての木の編集距離

木の編集距離は、2つの木の間距離(似ていない度合い)を表す尺度としてもっともよく使われている概念であり、ノードの削除、挿入、ラベルの書き換えの3つの編集操作によって、1つの木をもう一方の木に変換するための編集コストとして定義されている。各編集コストを1としたときの木の編集距離は、Tai 写像の概念を用いて、下記のように表現することができる。

木  $T_1$  から木  $T_2$  への Tai 写像  $M$  について、 $V_D = V(T_1) \setminus \{x | (x, y) \in M\}$ 、 $V_I = V(T_2) \setminus \{y | (x, y) \in M\}$  とおく。Tai 写像  $M$  のコスト  $d(M)$  を次のように定義する。

$$d(M) = \sum_{(x,y) \in M} (1 - \delta(\ell(x), \ell(y))) + |V_D| + |V_I|$$

ここで、 $\delta(x, y)$  は Kronecker のデルタとする。更に、 $\mathcal{M}(T_1, T_2)$

を、木  $T_1$  から  $T_2$  への全ての可能な木写像の集合とすると、 $T_1$  と  $T_2$  の編集距離は、次のように表すことができる。

$$D(T_1, T_2) = \min_{M \in \mathcal{M}(T_1, T_2)} \{d(M)\}$$

つまり、木の編集距離は、 $\mathcal{M}(T_1, T_2)$  の中から最小コストの木写像を見つける最適化問題である。

### 3.3 木写像のクラス階層

これまでに、編集操作(ノードの削除、挿入、ラベルの書き換え)の実行に制約を設けることで、特定の領域への適用や計算量の抑制を可能とする種々の編集距離が提案されている。これらの編集距離は、 $\mathcal{M}(T_1, T_2)$  の定義を狭めることにより、Tai 写像の場合と同じ式で定義することができる。各々の編集距離に対応する  $\mathcal{M}(T_1, T_2)$  の定義は、Tai 写像の集合の部分集合となる木写像のクラスを定めるが、その多くは融合可能写像、分離写像、位相保存写像、部分木同型写像のクラスのいずれかに帰着されることが知られている [4]。以下では、Tai 写像  $M$  の任意の要素  $(x_1, x_2), (y_1, y_2), (z_1, z_2) \in M$  に対して成り立つべき性質として、これらの木写像のクラスを定義する。また、図 1 でそれぞれの木写像クラスのイメージを図示するが、木  $T_1$  と  $T_2$  間の木写像によるノードの対応を破線、木写像に関わる木の構造を太線、木写像に含まれるノードを黒丸で表す。

表 1 木写像の最適化問題の時間計算量

木写像	順序付き木	順序無し木
Tai	$O(n^3 \log n)$	MAX-SNP 困難
融合可能	$O(n^2 d^2)$	MAX-SNP 困難
分離	$O(n^2)$	$O(n^2 2d \log_2 2d)$
位相保存	$O(n^2)$	$O(n^2 2d \log_2 2d)$
部分木同型	$O(n^2)$	$O(n^{2.5} / \log n)$

b) 融合可能写像 [4]

$$x_1 \sim y_1 < x_1 \sim z_1 \Rightarrow y_2 \sim z_2 = x_2 \sim z_2 \quad (\text{注1})$$

それぞれの木の先祖子孫の関係や左右を保ったまま、木写像に含まれるノード対を重ね合わせて、1つの木に融合するための必要十分条件を与える木写像のクラス。

c) 分離写像 [7] (注2)

$$z_1 < x_1 \sim y_1 \Leftrightarrow z_2 < x_2 \sim y_2$$

互いに素な木同士を比較して得られる木写像のクラスであり、この制約によって、最適化問題の時間計算量は、文字列と同等になる。

d) 位相保存写像 [8] (注3)

$$x_1 \sim y_1 = x_1 \sim z_1 \Leftrightarrow x_2 \sim y_2 = x_2 \sim z_2$$

図 1(d) からわかるように、木の位相構造が保存されるような木写像のクラスである。

e) 部分木同型写像

任意の  $(x_1 \sim y_1, x_2 \sim y_2) \in M$  について、

$$\forall w_1 \exists w_2 [x_1 < w_1 < y_1 \Rightarrow (w_1, w_2) \in M]$$

$$\forall w_2 \exists w_1 [x_2 < w_2 < y_2 \Rightarrow (w_1, w_2) \in M]$$

木の部分同型の成す木写像のクラスである。

これらの木写像が成す構造 (図 1 においては、太線の成すグラフ構造) は、部分木同型写像を除いて、一般に木構造のみでなく、森となる。

木写像のクラスは、図 1(f) に示すような階層関係が成り立つ。

表 1 にクラス階層と木の編集距離としての最適化問題の時間計算量の関係を示す。ここで、 $n$  は木のノード数、 $d$  は 2 つの木の中で最大の子供の数を表す。子ノードに順番がない順序無し木の計算量では、融合可能写像と分離写像の間に大きな溝があることがわかる。

### 3.4 畳み込みカーネルに基づく木カーネル

Hausser は、離散構造をもったデータは、その構造の部分構造により再帰的に特徴付けられるという考えに基づき、離散構造に対するカーネル関数設計の一般的な枠組みである畳み込みカーネルを提案した [1]。畳み込みカーネルは、2 つの構造データ  $T_1$  と  $T_2$  が与えられたとき、

$$K(T_1, T_2) = \sum_{t_1 \subseteq T_1} \sum_{t_2 \subseteq T_2} K_t(t_1, t_2) \quad (1)$$

と定義される。ここで、 $t \subseteq T$  は、 $t$  が、 $T$  の部分構造であることをあらわし、 $K_t$  は、2 つの部分構造の間に定義されるカーネル関数であるとする。

Collins と Duffy は、構文解析木間の畳み込みカーネルを設計した [9]。Collins 等のカーネルでは、部分構造を取り出す時に、子ノードから順に、かつ、左側から順に取り出さなければならぬという厳しい規則を設けていた。即ち、 $x \in t \subseteq T$  において、 $x$  の子孫を部分構造  $t$  に含めるのならば子ノードから順に、また、子ノードを含めるのならば最も左側 (長子) から

(注1) : alignable-mapping, less constrained mapping と等価

(注2) : isolated-subtree mapping, constrained mapping, structure-preserving mapping, structure-respecting mapping と等価。

(注3) : more constrained mapping と等価。

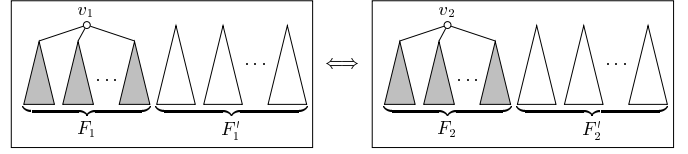


図 2 Tai 写像数え上げのための森の構造分解

順にとられなければならないとしていた。これに対し、鹿島らは、部分構造を取り出す際の規則をどこまで緩めることができるかという問題に取り組み、部分木同型写像、及び、位相保存写像によって定められる部分構造まで緩和できることを示した [5], [6]。即ち、鹿島等は、木  $t_1, t_2$  に対して、 $K_T(t_1, t_2)$  を、 $(\text{root}(t_1), \text{root}(t_2)) \in M$  を満足する部分木同型写像、或いは、位相保存写像の総数と定義する時、下式で定義されるカーネル関数が再帰式で表現され、従って、動的プログラミングにより効率的に評価することができることを示した。

$$K^*(T_1, T_2) = \sum_{v_1 \in V(T_1)} \sum_{v_2 \in V(T_2)} K_T(T_1(v_1), T_2(v_2))$$

一方、本論文では、鹿島等の結果を、以下の 2 方向で拡張する。

(1) 位相保存写像の制約を更に緩め、現在知られているより広い写像クラスである、分離写像、融合可能写像、及び、Tai 写像全てにおいても、カーネル関数を再帰式により表現できることを示す。

(2)  $K_T(t_1, t_2)$  において、根ノード同士を対応付けるという制約を緩和し、木写像 (部分木同型、位相同型、分離、融合可能、Tai) の像の共通祖先がそれぞれ  $t_1$  及び  $t_2$  の根ノードと一致するという制約に緩和する (注4)。

## 4. 木写像の数え上げ

この節では、木写像として Tai 写像、融合可能写像、分離写像、位相保存写像、部分木同型写像の 5 つを取り上げ、それぞれの木写像の数え上げ関数 ( $K(T_1, T_2)$ ) を再帰式で表現する。但し、数え上げにおいて重み付けを行わず、従って、ラベル間の類似度関数  $\sigma : \Sigma \times \Sigma \rightarrow \mathbb{R}$  は、Kronecker のデルタであるとする。

また、数え上げ関数  $K(T_1, T_2)$  は、 $K_T(t_1, t_2)$  を像の共通祖先が  $t_1$  及び  $t_2$  の根ノードと一致する木写像 (Tai、融合可能、分離、位相保存、部分木同型) の総数と定義した時の畳み込みカーネル関数  $K^*(T_1, T_2)$  に一致する。

以下では、最初に、Tai 写像の数え上げ関数の再帰式を与える。次いで、その他の木写像クラスに対する数え上げ関数を与えるが、これらのクラスに対しては数え上げのための共通の枠組みが存在するので、まずその枠組みについて述べ、続いて個別のクラスについて述べることにする。

### 4.1 Tai 写像の数え上げ関数

Tai 写像の定義は森間の Tai 写像に自然に拡張できるので、 $K(T_1, T_2)$  を  $K(F_1, F_2)$  に拡張することができる。

(注4) : この制約は  $K_T(t_1, t_2)$  間での数え上げの重複を避けるために必要である

Tai 写像の森カーネル  $K(F_1, F_2)$  は以下の式により再帰的に計算できる。図 2 に森の部分構造の対応を示す。

$$K(v_1(F_1) \circ F'_1, v_2(F_2) \circ F'_2) = \sigma(\ell(v_1), \ell(v_2)) \cdot (1 + K(F_1, F_2)) \cdot (1 + K(F'_1, F'_2)) \quad (2a)$$

$$+ K(v_1(F_1) \circ F'_1, F_2 \bullet F'_2) - K(F_1 \bullet F'_1, F_2 \bullet F'_2) \quad (2b)$$

$$+ K(F_1 \bullet F'_1, v_2(F_2) \circ F'_2) - K(F_1 \bullet F'_1, F_2 \bullet F'_2) \quad (2c)$$

$$+ K(F_1 \bullet F'_1, F_2 \bullet F'_2), \quad (2d)$$

$$K(F, \emptyset) = K(\emptyset, F) = 0,$$

この再帰式の意味を示す。

「森  $v_1(F_1) \circ F'_1$  と  $v_2(F_2) \circ F'_2$  の間の Tai 写像の総数」

= 「 $(v_1, v_2)$  を含む Tai 写像の数 (2a)」

+ 「 $y \neq v_2$  であるような  $(v_1, y)$  を含む Tai 写像の数 (2b)」

+ 「 $x \neq v_1$  であるような  $(x, v_2)$  を含む Tai 写像の数 (2c)」

+ 「 $v_1$  と  $v_2$  のいずれも含まないような Tai 写像の数 (2d)」

Zhang と Shasha による最適化問題と同様の技術 [10] を用いることにより、上記の再帰式を動的プログラミングにより、時間計算量  $O(|T_1| \cdot |T_2| \cdot \min\{L_1, D_1\} \cdot \min\{L_2, D_2\})$  で評価することができる ( $L_i, D_i$  は、それぞれ、 $T_i$  の葉の数と最大の深さを表すものとする)。

上記の再帰式は、次の命題 1 による。

**命題 1.** Tai 写像は以下の性質を満たす。

(1)  $F'_1 \subseteq F_1, F'_2 \subseteq F_2$  とする時、 $F'_1, F'_2$  間の木写像  $M$  に関して、(a) と (b) は互いに同値である。

(a)  $M$  は  $F_1, F_2$  間の Tai 写像である。

(b)  $M$  は  $F'_1, F'_2$  間の Tai 写像である。

(2)  $M$  を  $F_1$  から  $F_2$  への木写像、 $M'$  を  $F'_1$  から  $F'_2$  への木写像とすると、(a) と (b) は互いに同値である。

(a)  $M$  及び  $M'$  はそれぞれ Tai 写像である。

(b)  $M \cup M'$  は  $F_1 \bullet F'_1$  から  $F_2 \bullet F'_2$  への Tai 写像である。

(3)  $F_i$  を木  $T_i$  であるとする時、 $T_1 \setminus \{\text{root}(T_1)\}, T_2 \setminus \{\text{root}(T_2)\}$  間の木写像  $M$  に対して、(a) と (b) は互いに同値である。

(a)  $M$  は Tai 写像である。

(b)  $M \cup \{(\text{root}(T_1), \text{root}(T_2))\}$  は Tai 写像である。

## 4.2 数え上げ関数の共通の枠組み

### 4.2.1 準備

以下では、木写像に関する性質  $\mathcal{P}$  とは、融合可能、分離、位相保存、部分木同型のいずれかであるとする。

まず、木間の木写像の定義を森間の木写像に拡張する。Tai 写像の場合はこの拡張は自然であったが、性質  $\mathcal{P}$  の場合は、異なる木に属するノード  $x, y$  に対して  $x \sim y$  が定義されないことから、拡張には若干の工夫が必要である。本稿では、森  $F_i$  (木である場合も含む) に対するカーネル関数  $K(F_1, F_2)$  を以下のように定義する。

**定義 2.** 最初に、 $\bar{F}_i$  を次のように定義する。 $F_i$  が木である場合、 $\bar{F}_i = F_i$  とする。 $F_i$  が複数の木から構成される森である

場合、仮想的なルート  $v_i$  に対して  $\bar{F}_i = v_i(F_i)$  とする。 $F_1$  から  $F_2$  への写像  $M$  が性質  $\mathcal{P}$  を満たす木写像であるとは、 $M$  を  $\bar{F}_1, \bar{F}_2$  への木写像と自然にみなした時、性質  $\mathcal{P}$  を満足することと定義する。

この定義により、数え上げ関数も、 $K(T_1, T_2)$  から  $K(F_1, F_2)$  に拡張することができる。

拡張された数え上げ関数は、性質  $\mathcal{P}$  の種類に応じて 4 種存在するが、以下では表記上それらを区別せず、統一的に扱い、それぞれの節におけるコンテキストに従って解釈するものとする。

性質  $\mathcal{P}$  では、命題 1 に対応する性質は、以下の複数の命題により与えられる。

**命題 2.**  $F'_i \subseteq F_i$  は以下のいずれかであるとする。

- $F_i = F'_i$
- $F_i = (\text{root}(F_i)(F'_i))$
- $F_i = G \bullet F'_i \bullet H$  ( $G, H$  は順序付森)

この時、 $F'_1$  から  $F'_2$  への木写像  $M$  に対して、(1) と (2) は互いに同値である。

(1)  $M$  は  $F'_1$  から  $F'_2$  への木写像として性質  $\mathcal{P}$  を満足する。

(2)  $M$  は  $F_1$  から  $F_2$  への木写像として性質  $\mathcal{P}$  を満足する。

**命題 3.**  $M$  を  $F_1$  から  $F_2$  への木写像、 $M'$  を  $F'_1$  から  $F'_2$  への木写像とすると、(1) と (2) は互いに同値である。

(1)  $M$  及び  $M'$  はそれぞれ性質  $\mathcal{P}$  を満足する。

(2)  $M \cup M'$  は  $F_1 \bullet F'_1$  から  $F_2 \bullet F'_2$  への木写像として性質  $\mathcal{P}$  を満足する。

**命題 4.** Tai 写像  $M$  に対して、(1) と (2) は互いに同値である。

(1)  $M$  は融合可能 (分離) 写像である。

(2)  $\mathcal{L}(M)$  は融合可能 (分離) 写像である。

但し、 $\mathcal{L}(M) = \{(x, y) \in M \mid \exists (z, u) \in M [z < x]\}$  とする。

**命題 5.**  $F_1 = v_1(\prod_{i=1}^m T_i^1), F_2 = v_2(\prod_{i=2}^n T_i^2)$  とし、 $M$  を  $\prod_{i=1}^m T_i^1$  から  $\prod_{j=1}^n T_j^2$  への木写像とする時、次の (1) と (2) は互いに同値である。

(1)  $M \cup \{(v_1, v_2)\}$  は  $F_1$  から  $F_2$  への位相保存写像である。

(2)  $M$  は位相保存であり、かつ、 $M \cap (T_i^1 \times T_j^2) \neq \emptyset$  ならば下式が成り立つ。

$$\forall k \neq j \forall \ell \neq i [M \cap (T_i^1 \times T_k^2) = M \cap (T_\ell^1 \times T_j^2) = \emptyset]$$

### 4.2.2 数え上げ関数の雛形

$K(F_1, F_2)$  の再帰式を与えるに当たって、命題 2 により、下式を性質  $\mathcal{P}$  によらない共通の雛形として利用することができる。但し、 $T_1, T_2$  は空でない任意の木、 $F_1, F_2, F'_1, F'_2$  は任意の森であるとする。

$$K(T_1 \circ F_1, T_2 \circ F_2) = K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2) \quad (3a)$$

$$+ K(F_1, F_2) \quad (3b)$$

$$K(v(F_1), T_2 \circ F_2) = K(v(F_1), T_2) - K(F_1, T_2) \quad (4a)$$

$$+ K(v(F_1), F_2) - K(F_1, F_2) \quad (4b)$$

$$+ K(F_1, T_2 \circ F_2) \quad (4c)$$

$$K(v_1(F_1), v_2(F_2)) = \sigma(\ell(v_1), \ell(v_2)) \cdot (1 + K'_{\mathcal{P}}(F_1, F_2)) \quad (5a)$$

$$+ K(v_1(F_1), F_2) - K(F_1, F_2) \quad (5b)$$

$$+ K(F_1, v_2(F_2)) - K(F_1, F_2) \quad (5c)$$

$$+ K(F_1, F_2) \quad (5d)$$

$$K(F, \emptyset) = K(\emptyset, F) = 0$$

上記の再帰式の各項の意味を以下に与える。

(3a)  $K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  は、性質  $\mathcal{P}$  を満足する  $T_1 \circ F_1, T_2 \circ F_2$  間の木写像であり、かつ、 $T_1$  或いは  $T_2$  のノードを含むものの総数。

(3b) 性質  $\mathcal{P}$  を満足する  $T_1 \circ F_1, T_2 \circ F_2$  間の木写像であり、かつ、 $T_1$  及び  $T_2$  のノードを含まないものの総数。

(4) 性質  $\mathcal{P}$  を満足する  $v(F_1), T_2 \circ F_2$  間の木写像であり、かつ、以下のそれぞれを満足する木写像の総数。

(4a)  $v$  の対応が  $T_2$  中に存在する。

(4b)  $v$  の対応が  $F_2$  中に存在する。

(4c)  $v$  を含まない。

(5) 性質  $\mathcal{P}$  を満足する  $v_1(F'_1), v_2(F'_2)$  間の木写像であり、かつ、以下のそれぞれを満足する木写像の総数。

(5a)  $(v_1, v_2)$  を含む。  $K'_{\mathcal{P}}(F'_1, F'_2)$  は、 $F'_1, F'_2$  間の木写像  $M$  であり、かつ、 $M \cup \{(v_1, v_2)\}$  が性質  $\mathcal{P}$  を満足するものの総数となる。

(5b)  $v_1$  の対応が  $F'_2$  中に存在する。

(5c)  $v_2$  の対応が  $F'_1$  中に存在する。

(5d)  $v_1, v_2$  のいずれも含まない。

$K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  及び  $K'_{\mathcal{P}}(F'_1, F'_2)$  の評価式は性質  $\mathcal{P}$  に依存するので、以下では、融合可能写像、分離写像、位相保存写像、及び、部分木同型写像のそれぞれについて評価式を与える。

#### 4.2.3 部分木同型写像

$K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  の評価

$$K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2) = K(T_1, T_2) + K(T_1, F_2) + K(F_1, T_2)$$

$K'_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  の評価

以下の再帰式で計算される。

$$K'_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2) = K'_{\mathcal{P}}(T_1, T_2) \cdot (1 + K'_{\mathcal{P}}(F_1, F_2))$$

$$+ K'_{\mathcal{P}}(F_1, T_2 \circ F_2)$$

$$+ K'_{\mathcal{P}}(T_1 \circ F_1, F_2)$$

$$- K'_{\mathcal{P}}(F_1, F_2)$$

$$K'_{\mathcal{P}}(v_1(F_1), v_2(F_2)) = \sigma(\ell(v_1), \ell(v_2)) \cdot (1 + K'_{\mathcal{P}}(F_1, F_2))$$

$$K'_{\mathcal{P}}(F, \emptyset) = K'_{\mathcal{P}}(\emptyset, F) = 0$$

#### 4.2.4 位相保存写像

$\bar{K}(F_1, F_2)$  の評価

分離写像及び位相保存写像に対して、 $\bar{K}(F_1, F_2)$  を以下のよう  
に与える。

**定義 3.** 仮想的な根ノード  $v_1$  及び  $v_2$  に対し、 $\bar{F}_i = v_i(F_i)$  とする。 $\bar{K}(F_1, F_2)$  は、 $F_1$  から  $F_2$  への木写像で、以下の条件を満足するものの総数であると定義する。

(1)  $M$  は  $\bar{F}_1$  から  $\bar{F}_2$  への分離 (位相保存) 写像である。

(2) 任意の  $(x_1, y_1), (x_2, y_2) \in M$  において、 $x_1$  が  $y_1$  が ( $F_1$  に属する) 共通の根ノードの下にあることと、 $y_1$  と  $y_2$  が ( $F_2$  に属する) 共通の根ノードの下にあることは同値である。

命題 2 及び命題 3 により、 $\bar{K}(T_1 \circ F_1, T_2 \circ F_2)$  は、以下のよう  
に評価できる。

$$\bar{K}(T_1 \circ F_1, T_2 \circ F_2) = K(T_1, T_2)(1 + \bar{K}(F_1, F_2))$$

$$+ \bar{K}(F_1, T_2 \circ F_2) - \bar{K}(F_1, F_2)$$

$$+ \bar{K}(T_1 \circ F_1, F_2) - \bar{K}(F_1, F_2)$$

$$+ \bar{K}(F_1, F_2)$$

$K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  の評価

**命題 6.** 以下の (1) と (2) は互いに同値である。

(1)  $M$  は位相保存写像である。

(2) 以下が成立する。

(a) 任意の  $i = 1, \dots, m$  に対して、 $M_i^1$  は位相保存写像

(b) 任意の  $j = 1, \dots, n$  に対して、 $M_j^2$  は位相保存写像

(c)  $M_i^1 \cap M_j^2 \neq \emptyset$  ならば、以下のいずれかが成立

i.  $M = M_i^1$

ii.  $M = M_j^2$

iii.  $\forall k \neq j \forall \ell \neq i [M_i^1 \cap M_k^2 = M_\ell^1 \cap M_j^2 = \emptyset]$

命題 2、命題 3、及び、命題 6 により、以下の  $K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  の評価を得る。

$$K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2) = K(T_1, T_2)(1 + \bar{K}(F_1, F_2))$$

$$+ K(T_1 \circ F_1, T_2) - K(T_1, T_2)$$

$$+ K(T_1, T_2 \circ F_2) - K(T_1, T_2)$$

$K'_{\mathcal{P}}(F'_1, F'_2)$  の評価

命題 5 により、以下の評価式を得る。

$$K'_{\mathcal{P}}(F'_1, F'_2) = \bar{K}(F_1, F_2)$$

#### 4.2.5 分離写像

$K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  の評価

**命題 7.** 以下の (1) と (2) は互いに同値である。

(1)  $M$  は分離写像である。

(2) 以下が成立する。

(a) 任意の  $i = 1, \dots, m$  に対して、 $M_i^1$  は分離写像

(b) 任意の  $j = 1, \dots, n$  に対して、 $M_j^2$  は分離写像

(c)  $M_i^1 \cap M_j^2 \neq \emptyset$  ならば、以下のいずれかが成立

i.  $M = M_i^1$

ii.  $M = M_j^2$

iii.  $\forall k \neq j \forall \ell \neq i [M_i^1 \cap M_k^2 = M_\ell^1 \cap M_j^2 = \emptyset]$

命題 2、命題 3、及び、命題 7 により、以下の  $K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2)$  の評価を得る。

$$K_{\mathcal{P}}(T_1 \circ F_1, T_2 \circ F_2) = K(T_1, T_2)(1 + \bar{K}(F_1, F_2))$$

$$+ K(T_1 \circ F_1, T_2) - K(T_1, T_2)$$

$$+ K(T_1, T_2 \circ F_2) - K(T_1, T_2)$$

$K'_P(F'_1, F'_2)$  の評価 命題 4 により、 $K'_P(F'_1, F'_2)$  は以下のよう  
に評価される。

$$K'_P(F'_1, F'_2) = K(F'_1, F'_2)$$

#### 4.2.6 融合可能写像

$K_P(T_1 \circ F_1, T_2 \circ F_2)$  の評価

$T_i^1, T_j^2$  を木とし、 $\prod_{i=1}^m T_i^1, \prod_{j=2}^n T_j^2$  を森とする。 $1 \leq a \leq b \leq m$  及び  $1 \leq c \leq d \leq n$  に対し、 $\kappa_{c,d}^{a,b}$  を以下のように定める。

$$\kappa_{c,d}^{a,b} = K\left(\prod_{i=a}^b T_i^1, \prod_{j=c}^d T_j^2\right)$$

更に、 $M$  を  $v_1(\prod_{i=1}^m T_i^1)$  から  $v_2(\prod_{j=1}^n T_j^2)$  への木写像とし、

$$M_i^1 = M \cap (T_i^1 \times v_2 \left( \prod_{j=1}^n T_j^2 \right)), \quad M_j^2 = M \cap (v_1 \left( \prod_{i=1}^m T_i^1 \right) \times T_j^2)$$

と表す時、次の性質が成立する。

**命題 8.** 以下の (1) と (2) は互いに同値である。

- (1)  $M$  は融合可能である。
- (2) 以下が成立する。
  - (a) 任意の  $i = 1, \dots, m$  に対して、 $M_i^1$  は融合可能である。
  - (b) 任意の  $j = 1, \dots, n$  に対して、 $M_j^2$  は融合可能である。
  - (c)  $M_i^1 \cap M_j^2 \neq \emptyset, M_i^1 \setminus M_j^2 \neq \emptyset, M_j^2 \setminus M_i^1 \neq \emptyset$  を同時に満足する  $i, j$  は存在しない。

命題 2、命題 3、及び、命題 8 により、 $K_P(T_1 \circ F_1, T_2 \circ F_2)$  は以下の再帰式によって表される。

$$\begin{aligned} K_P(T_1^1 \circ \prod_{i=2}^m T_i^1, T_1^2 \circ \prod_{j=2}^n T_j^2) &= \kappa_{1,1}^{1,1} (1 + \kappa_{2,n}^{2,m}) \\ &+ \sum_{j=2}^n \left\{ (\kappa_{1,j}^{1,1} - \kappa_{1,j-1}^{1,1}) (1 + \kappa_{j+1,n}^{2,m}) \right. \\ &\quad \left. + \sum_{i=2}^m (\kappa_{j,j}^{1,i} - \kappa_{j,j}^{1,i-1} - \kappa_{j,j}^{2,i} + \kappa_{j,j}^{2,i-1}) (1 + \kappa_{j+1,n}^{i+1,m}) \right\} \\ &+ \sum_{i=2}^m \left\{ (\kappa_{1,1}^{1,i} - \kappa_{1,1}^{1,i-1}) (1 + \kappa_{2,n}^{i+1,m}) \right. \\ &\quad \left. + \sum_{j=2}^m (\kappa_{1,j}^{i,i} - \kappa_{1,j-1}^{i,i} - \kappa_{2,j}^{i,i} + \kappa_{2,j-1}^{i,i}) (1 + \kappa_{j+1,n}^{i+1,m}) \right\} \end{aligned}$$

$K'_P(F'_1, F'_2)$  の評価

命題 4 により、 $K'_P(F'_1, F'_2)$  は以下のように評価される。

$$K'_P(F'_1, F'_2) = K(F'_1, F'_2)$$

### 5. 既存の木の畳み込みカーネルとの関係

$\sigma(l_1, l_2)$  を Kronecker のデルタ以外の関数として、 $K(T_1, T_2)$  を計算することにより、重み付けられた木写像の数のえ上げを得ることができる。

**定義 4.**  $M = \{(x_1, y_1), \dots, (x_N, y_N)\}$  を  $T_1$  から  $T_2$  への (Tai, 融合可能、分離、位相保存) 写像とする。 $M$  の重み  $w(M)$  を以下のように定義する。

$$w(M) = \prod_{i=1}^N \sigma(\ell(x_i), \ell(y_i))$$

**定理 9.**  $T_1$  から  $T_2$  への (Tai, 融合可能、分離、位相保存) 写像の集合を  $\mathcal{M}(T_1, T_2)$  と表す時、

$$K(T_1, T_2) = \sum_{M \in \mathcal{M}(T_1, T_2)} w(M)$$

が成り立つ。

このような重み付けのもとで、鹿島らによるラベル付き順序木カーネルと柔軟性を持つ木カーネル [6] は、それぞれ部分木同型写像と位相保存写像の数のえ上げ関数を用いて下記のように定義したものと等価であることが容易に示せる。

$$K^*(T_1, T_2) = \sum_{v_1 \in V(T_1)} \sum_{v_2 \in V(T_2)} \sigma(\ell(v_1), \ell(v_2)) \cdot (1 + K'_P(F_1, F_2))$$

### 6. カーネル関数であることの証明と特徴空間

この節では、関数  $\omega(l) : \Sigma \rightarrow \mathbb{R}$  が与えられて、 $\sigma(l_1, l_2)$  が、

$$\sigma(l_1, l_2) = \omega(l_1)\omega(l_2)\delta(l_1, l_2)$$

と与えられる場合の関数  $K(x, y)$  の特徴空間を具体的に示すことにより、これまでに示した木写像がカーネル関数であることを証明する。但し、 $\delta(l_1, l_2)$  は Kronecker のデルタとする。

まず、 $n$  をある自然数とおき、関数の集合  $\mathcal{F}_n$  を以下のように定める。

$$\mathcal{F}_n = \{f | (f : \Sigma \rightarrow \mathbb{N} \cup \{0\}) \wedge \sum_{l \in \Sigma} f(l) = n\}$$

このとき、任意のラベル無し木  $\tau$  と、 $\tau$  から  $n$  のノードを取り出し、ラベル付けしたノード集合  $V_\tau \subseteq V(\tau)$  の対  $(\tau, V_\tau)$  を考える。この  $V_\tau$  の要素に対しては、関数  $f \in \mathcal{F}_n$  により、記号  $l \in \Sigma$  でラベル付けされたノードがちょうど  $f(l)$  個になるようなあらゆる組合せを考える。

すなわち、ノード数が  $m \geq 1$  であるような全てのラベル無し木からなる集合を  $\mathcal{T}^m$  とおくと、次のような集合  $\mathcal{T}_f (f \in \mathcal{F}_n)$  を考える。

$$\begin{aligned} \mathcal{T}_f &= \bigcup_{m=n}^{\infty} \left\{ (\tau, V_\tau) | \tau \in \mathcal{T}^m \wedge V_\tau \subseteq V(\tau) \wedge |V_\tau| = n \right. \\ &\quad \left. \wedge \forall l \in \Sigma [ |\{x \in V_\tau | \ell(x) = l\}| = f(l) ] \right\} \end{aligned}$$

以下では、この基礎集合  $\mathcal{T}_f$  に条件を加えることにより特徴ベクトルを定義し、各々の木写像のクラスの数のえ上げ関数が、カーネル関数として解釈できることを示す。

#### 6.1 位相保存写像

$\text{TP}_f$  を、 $\mathcal{T}_f$  に次の条件を加えたものであると定義する。 $(\tau, V_\tau) \in \mathcal{T}_f$  において、 $\tau$  の任意のノード  $x \in V(\tau)$  に対し、 $V_\tau$  中の必ずしも異なるとは限らない 2 ノード  $(y, z) \in V_\tau \times V_\tau$  が存在して、 $x = y \sim z$  が成り立つ。

このとき、基底を次のように定める。

$$\mathcal{B}_N^{\text{TP}} = \bigcup_{n=1}^N \left( \bigcup_{f \in \mathcal{F}_n} \text{TP}_f \right)$$

この基底における、木  $T$  の特徴ベクトルを  $\phi(T)$  と表記する。

その  $(\tau, V_\tau) \in \text{TP}_f$  成分を  $\phi_{(\tau, V_\tau)}(T)$  と表記し、次のように定義する。

$$\phi_{(\tau, V_\tau)}(T) = \prod_{l \in \Sigma} \omega(l)^{f(l)} \psi_{(\tau, V_\tau)}(T)$$

ただし、 $\psi_{(\tau, V_\tau)}(T)$  は、 $\tau$  から  $T$  への位相保存写像  $M$  で、以下の条件を満足するものの総数と定義する。

- (1)  $(x, y) \in M \Rightarrow x \in V_\tau$
- (2)  $\forall x \in V_\tau \exists y \in V(T) [(x, y) \in M]$

**定理 10.**  $N \geq \max(|V(T_1)|, |V(T_2)|)$  に対して次式が成立する。

$$\begin{aligned} K(T_1, T_2) &= \phi(T_1) \cdot \phi(T_2) \\ &= \sum_{(\tau, V_\tau) \in \mathcal{B}_N^{\text{TP}}} \phi_{(\tau, V_\tau)}(T_1) \cdot \phi_{(\tau, V_\tau)}(T_2) \end{aligned}$$

## 6.2 分離写像

$\text{SEP}_f$  を、 $\mathcal{T}_f$  に次の条件を加えたものであると定義する。 $\tau$  中の任意のノード  $x \in V(\tau)$  に対して、次のいずれかの条件が成り立つような  $V_\tau$  中の異なる 2 ノード  $y, z \in V_\tau$  ( $y \neq z$ ) が存在する。

- (1)  $x = y \sim z$  が成り立つ。
  - (2)  $x = y$  かつ  $z \in \text{ch}(x)$  が成り立つ。
- このとき、基底を以下のように定める。

$$\mathcal{B}_N^{\text{SEP}} = \bigcup_{n=1}^N \left( \bigcup_{f \in \mathcal{F}_n} \text{SEP}_f \right)$$

この基底における、木  $T$  の特徴ベクトルを  $\phi(T)$  と表記する。その  $(\tau, V_\tau) \in \text{SEP}_f$  成分を  $\phi_{(\tau, V_\tau)}(T)$  と表記し、次のように定義する。

$$\phi_{(\tau, V_\tau)}(T) = \prod_{l \in \Sigma} \omega(l)^{f(l)} \psi_{(\tau, V_\tau)}(T)$$

ここで、 $\psi_{(\tau, V_\tau)}(T)$  の定義は、 $M$  が分離写像であることを除き、位相保存写像の場合と同じである。

**定理 11.**  $N \geq \max(|V(T_1)|, |V(T_2)|)$  に対して次式が成立する。

$$\begin{aligned} K(T_1, T_2) &= \phi(T_1) \cdot \phi(T_2) \\ &= \sum_{(\tau, V_\tau) \in \mathcal{B}_N^{\text{SEP}}} \phi_{(\tau, V_\tau)}(T_1) \cdot \phi_{(\tau, V_\tau)}(T_2) \end{aligned}$$

## 7. Tai 写像

$\text{TAI}_f$  を、 $\mathcal{T}_f$  において、 $V_\tau = V(\tau)$  または  $V_\tau = V(\tau) \setminus \{\text{root}(\tau)\}$  の条件を付加した集合であるとする。

基底を以下のように定める。

$$\mathcal{B}_N^{\text{TAI}} = \bigcup_{n=1}^N \left( \bigcup_{f \in \mathcal{F}_n} \text{TAI}_f \right)$$

この基底における、木  $T$  の特徴ベクトルを  $\phi(T)$  と表記する。その  $(\tau, V_\tau) \in \text{TAI}_f$  成分を  $\phi_{(\tau, V_\tau)}(T)$  と表記し、次のように定義する。

$$\phi_{(\tau, V_\tau)}(T) = \prod_{l \in \Sigma} \omega(l)^{f(l)} \psi_{(\tau, V_\tau)}(T)$$

ここで、 $\psi_{(\tau, V_\tau)}(T)$  の定義は、 $M$  が Tai 写像であることを除き、位相保存写像の場合と同じである。

**定理 12.**  $N \geq \max(|V(T_1)|, |V(T_2)|)$  に対して次式が成立する。

$$\begin{aligned} K(T_1, T_2) &= \phi(T_1) \cdot \phi(T_2) \\ &= \sum_{(\tau, V_\tau) \in \mathcal{B}_N^{\text{TAI}}} \phi_{(\tau, V_\tau)}(T_1) \cdot \phi_{(\tau, V_\tau)}(T_2) \end{aligned}$$

## 8. おわりに

本稿では、木写像という 2 つの木構造間の対応関係を数え上げる効率のよいアルゴリズムを示した。また、木写像の数え上げ関数が、カーネル関数としての要件を満たしていることを、その特徴空間を明らかにすることにより厳密に示した。この成果は、カーネル関数の設計の観点からは、鹿島らにより設計された既存の木カーネルを拡張し、木の量み込みカーネルの一般的な設計手法を示すものである。ただし、融合可能写像の数え上げ関数が、カーネル関数であるかどうかは未解決である。

今後の課題は、融合可能写像の数え上げ関数がカーネル関数であるかどうかを示すこと、および、我々が明確にした木カーネルの性質を用いて、実際に問題領域に応じたよりよい木カーネルの設計を行い、その有効性を示すことが挙げられる。また、今回は考慮しなかった順序無し木については、一般的な木写像の数え上げによるアプローチでは、計算困難性の問題があることがわかっており [6]、順序無し木のための高速で柔軟な木カーネルの設計が必要である。

## 文献

- [1] D. Haussler: "Convolution kernels on discrete structures", UCSC-CRL 99-10, Dept. of Computer Science, University of California at Santa Cruz (1999).
- [2] T. Gärtner: "A survey of kernels for structured data.", SIGKDD Explorations, **5**, 1, pp. 49–58 (2003).
- [3] K.-C. Tai: "The tree-to-tree correction problem", Journal of ACM, **26**, 3, pp. 422–433 (1979).
- [4] T. Kuboyama, K. Shin, T. Miyahara and H. Yasuda: "A theoretical analysis of alignment and edit problems for trees", The 9th Italian Conference of Theoretical Computer Science, LNCS 3701, pp. 323–337 (2005).
- [5] H. Kashima and T. Koyanagi: "Kernels for semi-structured data.", the 9th International Conference on Machine Learning (ICML 2002), pp. 291–298 (2002).
- [6] 鹿島久嗣, 坂本比呂志, 小柳光生: "木構造データに対するカーネル関数の設計と解析", 人工知能学会論文誌, **21**, 1 (2006).
- [7] K. Zhang: "Algorithms for the constrained editing distance between ordered labeled trees and related problems", Pattern Recognition, **28**, 3, pp. 463–474 (1995).
- [8] T. Kuboyama and K. Shin: "Class identification for tree mappings", 情報処理学会 アルゴリズム研究会研究報告, AL-102-2, pp. 9–16 (2005).
- [9] M. Collins and N. Duffy: "Convolution kernels for natural language", Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001], MIT Press, pp. 625–632 (2001).
- [10] K. Zhang and D. Shasha: "Simple fast algorithms for the editing distance between trees and related problems", SIAM Journal on Computing, **18**, 6, pp. 1245–1262 (1989).