

# 多項式時間PAC学習可能な木幅定数形式グラフ体系 について

著者	正代 隆義
雑誌名	九州国際大学教養研究
巻	21
号	1
ページ	1-26
発行年	2014-12
URL	<a href="http://id.nii.ac.jp/1265/00000489/">http://id.nii.ac.jp/1265/00000489/</a>

# 多項式時間 PAC 学習可能な 木幅定数形式グラフ体系について

正 代 隆 義

## 1 はじめに

インターネットの超高速化や安価な大容量データデバイスの普及は、巨大データ（ビッグデータ）を効率よく解析するための基盤技術の研究開発に拍車をかけている。刻一刻と蓄積されつつあるデータの中には、テキストデータのような文字列だけからなるデータだけでなく、離散的な構造を持ったデータも多くある。離散的な構造を表現する形式としてグラフがある。グラフはオブジェクトを頂点、2つのオブジェクト間の関係を辺で表現した現実世界の数学的モデルである。例えば、薬理学の分野では化学化合物の分子構造を解明するために、原子を頂点、原子間の化学結合を辺としたグラフに対して、データマイニングの研究が行われている。また、社会学におけるソーシャルネットワークの解析や、ウェブページ間のリンク構造を表すウェブグラフを利用したウェブコミュニティの抽出など、グラフとして表現したデータを直接扱うこと、及びその技術の研究が急速に行われてきている。このようにグラフを対象とするデータマイニングをグラフマイニングとよぶ。

頻出部分グラフマイニングは、このようなグラフマイニングにおける課題として、熱心に研究されているもののうちのひとつである。これはグラフデータに頻出するローカルな共通性を抽出するものであり、Yan ら [23] の研究をはじめとして多くの研究成果がある。しかし、グラフデータを扱う計算問題の中には、一般のグラフを対象とした場合、探索すべき解の個数が組合せ爆発を起こして天文学的數字になるという厄介な問題点をもつものがある。実際、一般のグラフを扱う頻出部分グラフマイニングもその

ひとつである。そのような組合せ爆発を避けるためのアプローチとして、Horváth ら [9] はグラフ理論でよく知られている外平面的グラフクラスに着目し、効率的な頻出連結部分グラフマイニングアルゴリズムを提案した。また、Yamasaki ら [21, 22] は、外平面的グラフクラスに基づく項グラフパターンを定義し、計算論的学習理論の観点から、効率の良い機械学習アルゴリズムを提案した。項グラフパターンは、複数の頻出部分グラフがどのように配置されているか等のグローバルな構造を表現することが出来るグラフパターンである。

本論文では、形式グラフ体系 (Formal Graph System, FGS) により設計されたグラフパターンクラスの多項式時間学習可能性について議論する。形式グラフ体系は、Uchida ら [19] が導入したグラフを対象とする一種の論理プログラム [10] である。形式グラフ体系によって生成される言語を形式グラフ体系言語と呼ぶ。木、順序木、外平面的グラフのクラスは形式グラフ体系言語である。このことから、形式グラフ体系を用いることで今までのグラフパターン設計手法を統一的に扱うことができると考えられる。それに加えて、形式グラフ体系の論理プログラムのな特徴を用いることで、多項式時間学習可能な形式グラフ体系言語クラスを明らかにすることが期待できる。

計算論的学習理論は、ある概念を学習することが本質的にどれほどの計算時間やメモリなどの計算資源を必要とするかという問題を数学的に研究する分野である [15]。Valiant [20] は、この問題を扱う理論の枠組として、PAC 学習 (Probably Approximately Correctly Learning) を提案した。これは、学習アルゴリズムの出す仮説に誤差を許しており、たとえ目標とする概念と完全には一致しなくても、そのようなことが稀にしか起こらないのであれば、それを良い仮説とみなす。多項式時間 PAC 学習可能な言語クラスとして、基本形式体系 (Elementary Formal System, EFS) 言語のサブクラスがある。基本形式体系は、形式グラフ体系がグラフを直接扱うことに対して、文字列を直接扱うことができる一種の論理プログラムである。Miyano ら [11] は、パラメータ化された遺伝的基本形式体系の多項式時間 PAC 学習可能なクラスを示した。基本形式体系は、PAC 学習モデルによる結果以外にも、帰納推論モデル [2]、質問学習モデル [16] による学習可能性に關す

る結果があり、計算論的学習理論と馴染みが深い。

本論文では、変数独立な遺伝的形式グラフ体系 (Variable-independent Hereditary FGS, VH-FGS) により生成可能な形式グラフ体系言語のサブクラスを扱う。VH-FGS の階層を  $VH-FGS(m, k, t, r, w, d)$  で表現する。ここで、 $m, k, t, r, w, d$  はそれぞれ確定節 (本論文ではグラフ書き換え規則とよぶ) の個数、ヘッドに現れる変数の出現回数、ボディのアトムの個数、述語記号の引数の個数、変数のランクの最大値、項グラフパタンの頂点の最大次数である。また  $VH-FGS^-(m, k, t, r, w, d)$  を、 $VH-FGS(m, k, t, r, w, d)$  のサブクラスで、項グラフパタンの木幅の最大数が  $w$  で、項グラフパタンの異なる変数が頂点を共有しないという2つの性質を満たす形式グラフ体系で生成可能な形式グラフ体系言語の全体とする。本論文では、 $VH-FGS^-(m, k, t, r, w, d)$  が、もし  $m, k, t, r, w, d$  が定数ならば、多項式時間 PAC 学習可能であることを証明する。

## 2 準備

### 2.1 項グラフパタン

$\Sigma$  と  $\Lambda$  を有限アルファベットとする。また、 $X$  を無限個の記号からなるアルファベットとする。 $X$  の記号を変数とよぶ。 $X$  中の記号  $x$  はランクとよばれる非負整数を持つ。変数  $x \in X$  のランクを  $rank(x)$  で表す。また、アルファベット  $\Sigma, \Lambda, X$  の間には、 $\Sigma \cap X = \emptyset$  および  $\Lambda \cap X = \emptyset$  が成り立つとする。

集合または順序付きリスト  $S$  に対して、 $|S|$  を  $S$  に含まれる要素の個数とする。

**定義 1** アルファベット  $\langle \Sigma, \Lambda, X \rangle$  上の項グラフパタン、または項グラフとは、次の4つの条件を満たす7つ組  $g = (V, E, \varphi, \psi, H, \lambda, ports)$  である。

1.  $(V, E)$  は頂点ラベルおよび辺ラベル付き有向グラフまたは単純無向グラフである。 $\varphi$  は頂点ラベル付け関数  $\varphi: V \rightarrow \Sigma$  であり、 $\psi$  は辺

ラベル付け関数  $\psi: E \rightarrow \Lambda$  である.

2.  $H$  は空でない超辺の集合である. すなわち,  $H \subseteq 2^V \setminus \{\emptyset\}$  である.
3.  $\lambda$  は超辺への変数ラベル付け関数  $\lambda: H \rightarrow X$  で,  $\text{rank}(\lambda(h)) = |h|$  を満たす関数である. 超辺  $h \in H$  に対して,  $\text{rank}(\lambda(h))$  を超辺  $h$  のランクとよぶ.
4.  $\text{ports}$  は超辺  $h \in H$  に,  $h$  に含まれる頂点から成る順序付きリストを対応させる関数である. 具体的には,  $\text{ports}$  は関数  $\text{ports}: H \rightarrow V^*$  で, 各変数  $h \in H$  に対して,  $\text{ports}(h)$  は  $\text{rank}(\lambda(h))$  個の異なる頂点からなる順序付きリストである.  $\text{ports}(h)$  中の頂点を超辺  $h$  のポートとよぶ. 文脈より明らかなきとき, 以下では超辺のことを変数とよぶことがある.

超辺集合  $H$  が空である項グラフを基礎項グラフとよぶ. 基礎項グラフは, 通常のラベル付きグラフとみなすことができる.  $(\Sigma, \Lambda, X)$  上の項グラフ全体を  $\mathcal{G}(\Sigma, \Lambda, X)$  と記述する. また,  $(\Sigma, \Lambda, X)$  上の基礎項グラフ全体を  $\mathcal{G}(\Sigma, \Lambda)$  と記述する.

項グラフ  $g = (V, E, \varphi, \psi, H, \lambda, \text{ports})$  に対して, ラベル付け関数  $\varphi, \psi, \lambda$  の値域をそれぞれ  $\varphi(V), \psi(E), \lambda(H)$  と記述する. 本論文では, 異なる 2 つの項グラフ  $g_1 = (V_1, E_1, \varphi_1, \psi_1, H_1, \lambda_1, \text{ports}_1)$  と  $g_2 = (V_2, E_2, \varphi_2, \psi_2, H_2, \lambda_2, \text{ports}_2)$  に対して,  $\lambda_1(H_1) \cap \lambda_2(H_2) = \emptyset$  と仮定する.

定義 2 項グラフ  $g_1 = (V_1, E_1, \varphi_1, \psi_1, H_1, \lambda_1, \text{ports}_1)$  と項グラフ  $g_2 = (V_2, E_2, \varphi_2, \psi_2, H_2, \lambda_2, \text{ports}_2)$  が項グラフ同型, または同型であるとは, 次の 3 つの条件を満たすときをいう.

1.  $(V_1, E_1)$  と  $(V_2, E_2)$  は頂点ラベルおよび辺ラベル付きグラフとして同型である.  $\pi$  を  $(V_1, E_1)$  から  $(V_2, E_2)$  への同型写像  $\pi: V_1 \rightarrow V_2$  とする.
2.  $H_1$  から  $H_2$  への全単射  $\omega: H_1 \rightarrow H_2$  が存在し, 任意の超辺  $h \in H_1$  に対して,  $\text{ports}_2(\omega(h)) = \pi^*(\text{ports}_1(h))$  が成り立つ. ただし,

$\pi^*((v_1, \dots, v_m)) = (\pi(v_1), \dots, \pi(v_m))$  とする.

3. 超辺  $h, h' \in H_1$  に対して,  $\lambda_1(h) = \lambda_1(h')$  のとき, かつそのときに限り  $\lambda_2(\omega(h)) = \lambda_2(\omega(h'))$  である.

$g_1$  と  $g_2$  が同型であることを,  $g_1 \cong g_2$  と記述する.

$g = (V, E, \varphi, \psi, H, \lambda, ports)$  を項グラフとする.  $V', E', H'$  をそれぞれ  $V, E, H$  の部分集合とする. 頂点ラベル付け関数  $\varphi|V'$ , 辺ラベル付け関数を  $\psi|E'$ , 超辺ラベル付け関数を  $\lambda|H'$  を, それぞれの定義域を  $V', E', H'$  に制限したラベル付け関数  $\varphi, \psi, \lambda$  とする. また同様に, 関数  $ports|H'$  を, その定義域を  $H'$  に制限した関数  $ports$  とする.

定義 3 項グラフ  $g_1 = (V_1, E_1, \varphi_1, \psi_1, H_1, \lambda_1, ports_1)$  と項グラフ  $g_2 = (V_2, E_2, \varphi_2, \psi_2, H_2, \lambda_2, ports_2)$  に対して,  $g_1$  が  $g_2$  の部分項グラフであるとは,  $V_1 \subseteq V_2, E_1 \subseteq E_2, H_1 \subseteq H_2$  であって,  $\varphi_1 = \varphi_2|V_1, \psi_1 = \psi_2|E_1, \lambda_1 = \lambda_2|H_1, ports_1 = ports_2|H_1$  が成り立つときをいう.

定義 4  $g_1$  と  $g_2$  を項グラフとする.  $g_1$  が  $g_2$  に部分項グラフ同型であるとは,  $g_2$  の部分項グラフ  $g'_2$  が存在して,  $g_1$  と  $g'_2$  が同型であるときをいう.

定義 5 項グラフ  $g = (V, E, \varphi, \psi, H, \lambda, ports)$  に対して,  $E = \emptyset$  かつ  $H = \{V\}$  ならば,  $g$  を星項グラフとよぶ. 星項グラフはただ1つの変数から成る項グラフである.

定義 6 項グラフ  $g = (V, E, \varphi, \psi, H, \lambda, ports)$  が連結であるとは, 任意の2頂点  $u, v \in V$  に対して, 頂点列  $u = u_0, u_1, \dots, u_\ell = v$  が存在して, その列の隣り合った任意の2頂点  $u_{i-1}, u_i$  ( $1 \leq i \leq \ell$ ) が, 辺で結ばれているか, またはひとつの超辺にポートとして含まれているか, どちらかが成り立つときをいう.

## 2.2 形式グラフ体系

$g_1, \dots, g_n$  を  $\mathcal{G}(\Sigma, \Lambda, X)$  に属す  $n$  個の項グラフとする. 引数  $n$  の述語記号  $p$  に対して,  $p(g_1, \dots, g_n)$  をアトムとよぶ.  $A, B_1, \dots, B_m$  をアトムとする. ただし,  $m \geq 0$  である. グラフ書き換え規則とは,  $A \leftarrow B_1, \dots, B_m$  という形の確定節である. グラフ書き換え規則  $A \leftarrow B_1, \dots, B_m$  に対して, アトム  $A$  をヘッドとよび, 左矢印  $\leftarrow$  の左側のアトムをそのグラフ書き換え規則のボディとよぶ. 2つのアトム  $p(f_1, \dots, f_n)$  と  $p(g_1, \dots, g_n)$  に対して, 全ての  $i = 1, \dots, n$  について  $f_i \cong g_i$  であるならば,  $p(f_1, \dots, f_n) \cong p(g_1, \dots, g_n)$  と記述する. 2つのグラフ書き換え規則  $A_0 \leftarrow A_1, \dots, A_m$  と  $A'_0 \leftarrow A'_1, \dots, A'_m$  に対して, 全ての  $i = 0, \dots, m$  について  $A_i \cong A'_i$  であるならば,  $(A_0 \leftarrow A_1, \dots, A_m) \cong (A'_0 \leftarrow A'_1, \dots, A'_m)$  と記述する.

**定義 7** 形式グラフ体系とはグラフ書き換え規則の有限集合である.

**定義 8** 項グラフ  $f$  と  $f$  の頂点の順序付きリスト  $\sigma$  に対して, 形式  $[f, \sigma]$  を項グラフパタン断片, または項グラフ断片とよぶ.

**定義 9** 項グラフ  $g = (V_g, E_g, \varphi_g, \psi_g, H_g, \lambda_g, ports_g)$  の超辺  $h = (u_1, \dots, u_\ell) \in H_g$  に対して, 項グラフ  $f = (V_f, E_f, \varphi_f, \psi_f, H_f, \lambda_f, ports_f)$  の  $h$  への超辺置換を次のように定める. ここで,  $|V_f| \geq \ell$  とする.  $\sigma = (v_1, \dots, v_\ell)$  を  $V_f$  の頂点の順序付きリストとする. 項グラフ断片  $[f, \sigma]$  に対して,  $f$  上の超辺置換  $h \leftarrow [f, \sigma]$  を次の手続きとする.  $f' = (V_{f'}, E_{f'}, \varphi_{f'}, \psi_{f'}, H_{f'}, \lambda_{f'}, ports_{f'})$  を  $f$  と同型な項グラフとする. 任意の頂点  $v \in V_f$  に対して,  $v$  に対応する  $f'$  の頂点を  $v'$  と記述する. 超辺  $h$  を  $H_g$  から削除し,  $h$  のポート  $u_1, \dots, u_\ell$  を  $f'$  の頂点  $v'_1, \dots, v'_\ell$  と, この順で同一視して,  $f'$  を  $g$  に貼付ける. もし, この手続きによって, 多重辺が出来るならば, 対応する  $f'$  の辺を削除する. また,  $u_i$  の新しい頂点ラベルを  $\varphi_g(u_i)$  とする.

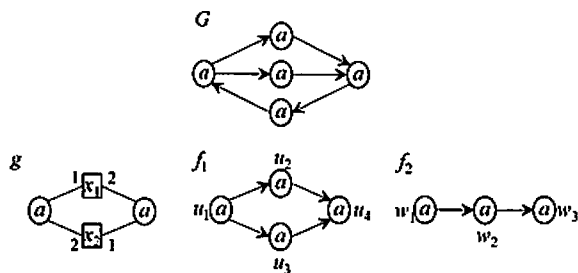


図1 代入: 有向グラフ  $G$  は項グラフ  $g$  に対して代入  $\theta = \{x_1 := [f_1, (u_1, u_4)], x_2 := [f_2, (w_3, w_1)]\}$  を適用することによって得られる。すなわち、 $G \cong g\theta$  である。

$\Upsilon = \{e_1 \leftarrow R_1, \dots, e_m \leftarrow R_m\}$  を  $f$  上の超辺置換の集合とする。  $f(\Upsilon)$  を  $\Upsilon$  の全ての超辺置換を同時に適用して得られる項グラフとする。

定義 10  $x$  を  $X$  の変数,  $[f, \sigma]$  を項グラフ断片とする。ただし,  $rank(x) = |\sigma|$  とする。このとき, 形式  $x := [f, \sigma]$  を  $x$  の束縛とよぶ。

定義 11 代入とは束縛の有限集合  $\theta = \{x_1 := [f_1, \sigma_1], \dots, x_n := [f_n, \sigma_n]\}$  である。ただし,  $x_1, \dots, x_n$  を互いに異なる  $X$  の変数とし, 各  $f_i$  は  $x_1, \dots, x_n$  でラベル付けされた超辺を持たないとする。

代入  $\theta$  とアトム  $p(g_1, \dots, g_n)$  に対して,  $p(g_1, \dots, g_n)\theta$  を  $p(g_1\theta, \dots, g_n\theta)$  と定める。グラフ書き換え規則  $A \leftarrow B_1, \dots, B_m$  に対して,  $(A \leftarrow B_1, \dots, B_m)\theta$  を  $A\theta \leftarrow B_1\theta, \dots, B_m\theta$  と定める。図1に, 項グラフと代入の例を与える。また, 図2に, 形式グラフ体系の例を与える。本論文では, 項グラフの超辺を, その超辺の変数を囲む正方形で表す。その超辺のポートを正方形から始まる線分のもう一方の端点で表し, 線分上にそのポートの超辺における順番を記述する。

項グラフ  $g_1, g_2$  と代入  $\theta$  に対して,  $g_1\theta \cong g_2\theta$  が成り立つとき, 代入  $\theta$  を  $g_1$  と  $g_2$  の単一化代入とよぶ。項グラフ  $g_1$  と  $g_2$  に対して, 代入  $\theta$  と  $\theta'$  が



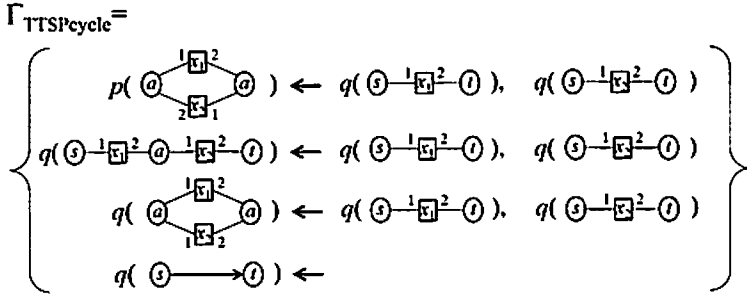


図 2 形式グラフ体系:  $\Gamma_{\text{TTSFCycle}}$

存在して、 $g_1 \cong g_2\theta$  かつ  $g_1\theta' \cong g_2$  が成り立つとき、 $g_1$  を  $g_2$  の変種とよぶ。アトムとグラフ書き換え規則に対しても、同様に単一化代入と変種を定める。ゴールとは、 $\leftarrow B_1, \dots, B_m$  ( $m \geq 0$ ) という形のグラフ書き換え規則である。もし  $m = 1$  ならば、そのゴール “ $\leftarrow B_1$ ” を単一ゴールとよぶ。また、もし  $m = 0$  ならば、空ゴールとよぶ。

グラフ書き換え規則  $C$  に対して、 $\text{var}(C)$  を  $C$  に現れる項グラフの変数全体とする。

定義 12  $\Gamma$  を形式グラフ体系とし、 $D$  をゴールとする。 $D$  からの導出演繹とは、次の条件を満たす 3 つ組の列  $(D_i, \theta_i, C_i)$  ( $i = 0, 1, \dots$ ) である。

1.  $D_i$  はゴール、 $\theta_i$  は代入、 $C_i$  は  $\Gamma$  のグラフ書き換え規則の変種である。
2.  $D_0$  は  $D$  である。
3.  $\text{var}(C_i) \cap \text{var}(C_j) = \emptyset$  ( $i \neq j$ ) であり、任意の  $i$  に対して、 $\text{var}(C_i) \cap \text{var}(D) = \emptyset$  である。
4. ゴールのボディからアトムをひとつ選ぶ手続き  $Q$  を仮定する。 $D_i$  をゴール  $\leftarrow A_1, \dots, A_k$  とし、 $C_i$  をグラフ書き換え規則  $A \leftarrow B_1, \dots, B_q$  とする。 $A_m$  ( $1 \leq m \leq k$ ) を  $Q$  によって選ばれた  $D_i$  のボディのアトムとする。このと

き,  $\theta_i$  は  $A$  と  $A_m$  の単一化代入であり,  $D_{i+1}$  はゴール  $\leftarrow A_1\theta_i, \dots, A_{m-1}\theta_i, B_1\theta_i, \dots, B_q\theta_i, A_{m+1}\theta_i, \dots, A_k\theta_i$  である.

反駁とは, 空ゴールで終わる有限長の導出演繹である.

定義 13  $F = \{(D_i, \theta_i, C_i)\}_{0 \leq i \leq k}$  を  $\Gamma$  の単一ゴールからの反駁とする.  $F$  の反駁木とは, 次のように定義される根付き木  $T_{k+1}$  である.

1. 各頂点は単一ゴールまたは空ゴールでラベル付けされている.
2. 根のラベルは単一ゴール  $D_0 = D$  である.
3. 各葉は空ゴールでラベル付けされている.
4.  $T_0$  は  $D$  でラベル付けされた頂点だけからなる根付き木である. 任意の  $i = 0, \dots, k$  に対して,  $T_{i+1}$  は,  $T_i$  に  $(D_i, \theta_i, C_i)$  を次のように適用して得られる根付き木である.  $D_i$  をゴール  $\leftarrow A_1^i, \dots, A_{n_i}^i$  とする.  $T_i$  は  $\leftarrow A_j^i$  ( $1 \leq j \leq n_i$ ) でラベル付けされた葉を持つと仮定する.  $C_i$  をグラフ書き換え規則  $A^i \leftarrow B_1^i, \dots, B_{q_i}^i$  とし,  $\theta_i$  を  $A^i$  と  $A_j^i$  の単一化代入とする. このとき,  $T_{i+1}$  は,  $\leftarrow A_j^i$  でラベル付けされた頂点に, その子供として,  $\leftarrow B_j^i\theta_i$  ( $1 \leq j \leq q_i$ ) でラベル付けされた  $q_i$  個の頂点を新しく接続した根付き木である.

反駁木の例を図 3 に与える.

定義 14  $\Gamma$  を形式グラフ体系とする. グラフ書き換え規則  $C$  に対する関係  $\Gamma \vdash C$  を次のように定める.

1. もし  $C \in \Gamma$  ならば,  $\Gamma \vdash C$  である.
2. もし  $\Gamma \vdash C$  ならば, 任意の代入  $\theta$  に対して,  $\Gamma \vdash C\theta$  である.
3. もし  $\Gamma \vdash A \leftarrow B_1, \dots, B_n$  かつ  $\Gamma \vdash B_i \leftarrow C_1, \dots, C_m$  ならば, 次が成り立つ.

$$\Gamma \vdash A \leftarrow B_1, \dots, B_{i-1}, C_1, \dots, C_m, B_{i+1}, \dots, B_n.$$

グラフ書き換え規則  $C$  に対して,  $\Gamma \vdash C$  であるとき,  $C$  は  $\Gamma$  から証明可能であるという.

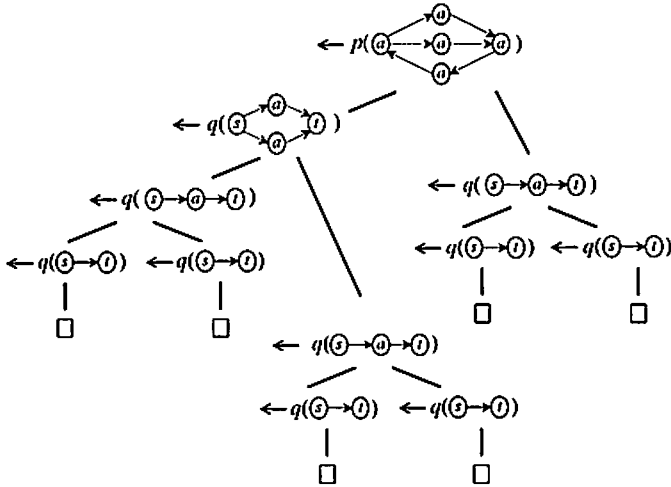


図3 根の単一ゴールから形式グラフ体系  $\Gamma_{TSP_{cyclic}}$ (図2) により得られる反駁木

以降, 引数の個数 1 の述語記号を単項述語記号とよぶ. 形式グラフ体系  $\Gamma$  と単項述語記号  $p$  に対して, 形式グラフ体系言語を次のように定める.

$$GL(\Gamma, p) = \{g \in \mathcal{G}(\Sigma, \Lambda) \mid \Gamma \vdash p(g) \leftarrow\}.$$

また形式グラフ体系言語クラスを次のように定める.

$$FGSL = \{GL(\Gamma, p) \mid \Gamma \text{ と } p \text{ はそれぞれ形式グラフ体系と単項述語記号}\}.$$

### 2.3 部分グラフの境界表現

$G = (V, E)$  を頂点ラベル付け関数  $\varphi$  と辺ラベル付け関数  $\psi$  を持つ頂点ラベルおよび辺ラベル付き有向グラフまたは単純無向グラフとする.  $E'$  を  $E$  の部分集合とする.  $G'$  を  $E'$  により誘導される  $G$  の辺誘導部分グラフとする. 本論文では辺誘導部分グラフを単に部分グラフとよぶ.  $\alpha = (V_\alpha, E_\alpha)$  を  $G$  の連結部分グラフとする. このとき,  $G \setminus \alpha$  を  $E \setminus E_\alpha$  によって誘導される  $G$  の部分グラフとする.

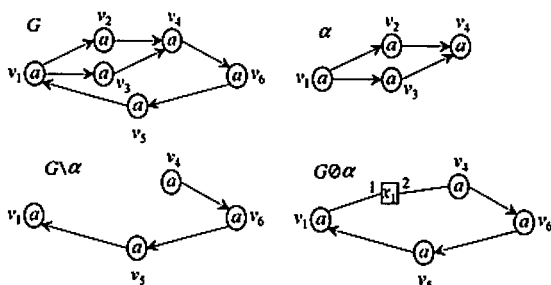


図4 連結グラフ  $G$  と  $G$  の連結部分グラフ  $\alpha$  に対して、部分グラフ  $G \setminus \alpha$  を与える。  $(\{v_1, v_4\}, \{(v_1, v_2), (v_1, v_3), (v_2, v_4), (v_3, v_4)\}, v_6, false)$  は  $\alpha$  の境界表現のひとつである。

定義 15  $\alpha = (V_\alpha, E_\alpha)$  を  $G = (V, E)$  の連結部分グラフとする。  $\alpha$  の境界頂点とは、  $G \setminus \alpha$  の辺の端点である  $\alpha$  の頂点である。  $\alpha$  の境界辺とは、境界頂点を端点とする  $\alpha$  の辺である。  $\alpha$  の境界表現とは、4 つ組  $(bv(\alpha), be(\alpha), m, IsIn(m, \alpha))$  である。各成分は次のことを表す。

- $bv(\alpha)$  は  $\alpha$  の境界頂点全体である。
- $be(\alpha)$  は  $\alpha$  の境界辺全体である。
- $m$  は  $G$  の任意の頂点である。この頂点を  $\alpha$  のマーカーとよぶ。
- $IsIn(m, \alpha)$  はマーカー  $m$  が  $\alpha$  に含まれるか否かを示す述語である。すなわち、もし  $m \in V_\alpha$  ならば、  $IsIn(m, \alpha) = true$  であり、そうでなければ  $IsIn(m, \alpha) = false$  である。

図4に境界表現の例を与える。部分グラフ  $\alpha$  の境界表現は  $\alpha$  を簡潔に示すために十分な表現である。  $\alpha$  と  $\alpha'$  を  $G$  の連結部分グラフとする。Chiangら [5] は、もし  $\alpha$  と  $\alpha'$  が同じ境界表現を持てば、  $\alpha \cong \alpha'$  であることを示した。

## 2.4 項グラフパタンの木幅

定義 16 項グラフ  $g = (V, E, \varphi, \psi, H, \lambda, ports)$  の木分解とは, 次の条件を満たす根付き木  $\mathcal{T} = (\mathcal{I}, \mathcal{F})$  である. 頂点  $i \in \mathcal{I}$  には  $V$  の部分集合  $V_i$ ,  $E$  の部分集合  $E_i$ ,  $H$  の部分集合  $H_i$  が対応する.

1. 頂点  $v \in V$  に対して,  $v \in V_i$  となる  $\mathcal{T}$  の頂点  $i \in \mathcal{I}$  が存在する.
2. 辺  $e = \{u, v\} \in E$  に対して,  $u, v \in V_i$  かつ  $e \in E_i$  となる  $\mathcal{T}$  の唯一の頂点  $i \in \mathcal{I}$  が存在する.
3. 超辺  $h = \{v_1, \dots, v_m\} \in H$  に対して,  $v_1, \dots, v_m \in V_i$  かつ  $h \in H_i$  となる  $\mathcal{T}$  の唯一の頂点  $i \in \mathcal{I}$  が存在する.
4. 頂点  $v \in V$  に対して,  $\{i \in \mathcal{I} \mid v \in V_i\}$  により誘導される  $\mathcal{T}$  の部分グラフは連結, すなわち部分木である.

$\mathcal{T} = (\mathcal{I}, \mathcal{F})$  の幅を  $\max_{i \in \mathcal{I}} |V_i| - 1$  と定める.  $g$  の木幅を,  $g$  の全ての木分解  $\mathcal{T}$  の幅のうち, 最小の幅と定める. 項グラフ断片  $[f, \sigma]$  の木分解とは,  $\sigma$  の全ての頂点を含む  $V_i$  ( $i \in \mathcal{I}$ ) が存在する  $f$  の木分解である. 項グラフの木分解は, その頂点  $i \in \mathcal{I}$  が次の3つのうちいずれかを満たすように変換することができる [6].

- 葉であり,  $V_i = \emptyset$  を満たす.
- $|E_i| + |H_i| = 1$  を満たし, ちょうど1つの子供を持つ.
- $|E_i| + |H_i| = 0$  を満たし, ちょうど2つの子供を持つ.

## 2.5 概念学習

$\Upsilon$  を有限アルファベットとする.  $\Upsilon^*$  の部分集合  $c$  を概念とよぶ. 概念  $c$  は,  $\Upsilon^*$  上の二値関数とみなすことができる. すなわち, 任意の  $w \in \Upsilon^*$  に対して,  $w \in c$  ならば1を,  $w \notin c$  ならば0を返す関数  $c: \Upsilon^* \rightarrow \{0, 1\}$  とみなせる. 概念クラスとは, 空でない概念の集合  $\mathcal{C} \subseteq 2^{\Upsilon^*}$  のことである.

概念を表現する有限アルファベットを  $\Delta$  とする. 概念クラス  $\mathcal{C}$  の表現と

は、関数  $R: C \rightarrow 2^{\Delta^*}$  で、任意の概念  $c$  に対して、 $R(c)$  は  $\Delta^*$  の空でない部分集合で、任意の異なる2つの概念  $c_1$  と  $c_2$  に対して、 $R(c_1) \cap R(c_2) = \emptyset$  が成り立つものをいう。各概念  $c \in C$  に対して、 $R(c)$  は  $c$  に対する名前の集合である。名前  $\nu \in R(c)$  の長さを  $\nu$  の文字列としての長さ  $|\nu|$  とする。概念  $c$  に対する名前の最小長を  $\ell_{\min}(c, R)$  とする。

本論文では、概念クラスとして形式グラフ体系言語クラス  $FGSL$  を扱う。また、概念クラス  $FGSL$  の表現として形式グラフ体系  $\Gamma$  と単項述語記号  $p$  のペア  $(\Gamma, p)$  全体を用いる。

例とは  $\Upsilon^* \times \{0, 1\}$  の要素  $\langle w, a \rangle$  のことをいう。概念  $c$  に対する例  $w \in \Upsilon^*$  とは  $\langle w, c(w) \rangle$  のことをいう。例の集合  $S \subseteq \Upsilon^* \times \{0, 1\}$  に対して、 $S_+ = \{w \mid \langle w, 1 \rangle \in S\}$ ,  $S_- = \{w \mid \langle w, 0 \rangle \in S\}$  と定める。概念  $c \in C$  が例の集合  $S$  に無矛盾であるとは、全ての  $w \in S_+$  に対して  $c(w) = 1$  であり、かつ全ての  $w \in S_-$  に対して  $c(w) = 0$  であるときをいう。

定義 17 概念クラス  $C$  が表現  $R$  で多項式時間 PAC 学習可能、または多項式時間学習可能であるとは、次の条件を満たすアルゴリズム  $A$  と多項式  $poly(\cdot, \cdot, \cdot, \cdot)$  が存在するときをいう。任意の概念  $c \in C$ 、任意の実数  $\epsilon, \delta$  ( $0 < \epsilon, \delta < 1$ )、任意の整数  $n \geq 0$ ,  $s \geq 1$ 、さらに  $\Upsilon^{[n]}$  上の任意の確率分布  $P$  に対して、

1.  $A$  は  $\epsilon, \delta, n, s$  を入力とする。(  $\epsilon$  を精度,  $\delta$  を信頼度,  $n$  を長さパラメータ,  $s$  を概念複雑さとよぶ。 )
2. EXAMPLE を  $\Upsilon^{[n]}$  上の確率分布  $P$  に従って  $c \in C$  の例をランダムに生成できる手続きとする。 $A$  は EXAMPLE を呼び出してもよい。
3.  $A$  は、次の条件を満たす概念  $h \in C$  の名前  $\nu \in R(h)$  を出力する。 $\ell_{\min}(c, R) \leq s$  を満たすとき、確率  $1 - \delta$  以上で、 $P(c \cup h - c \cap h) < \epsilon$  である。
4.  $A$  の計算時間は  $poly(\frac{1}{\epsilon}, \frac{1}{\delta}, n, s)$  で抑えられる。

定義 18 概念クラス  $C$  の表現  $R$  が多項式時間計算可能であるとは、決定性アルゴリズム  $B$  と多項式  $poly$  が存在して、次の (a) と (b) を満たすときを

いう。

- (a)  $B$  は入力として,  $w \in \Upsilon^*$  と  $\nu \in \Delta^*$  を受け取る。
- (b) もし, ある  $c \in C$  に対して,  $\nu \in R(c)$  であるならば,  $B$  は  $\text{poly}(|w| + |\nu|)$  時間で停止して,  $c(w)$  を出力する。

任意の例の集合  $S$  に対して,  $\ell_{\min}(S, R)$  を, 概念クラス  $C$  の表現  $R$  に含まれる表現のうち  $S$  と無矛盾な表現の最小長のこととする。

**定義 19**  $C$  を概念クラス,  $R$  をその表現,  $S \subseteq \Upsilon^* \times \{0, 1\}$  を例の有限集合とする.  $S$  を入力とする決定性アルゴリズム  $A$  が適合であるとは,  $S$  に無矛盾な概念  $c \in C$  が存在すれば, その表現  $\nu \in R(c)$  を出力するときをいう.  $A$  が多項式時間適合であるとは,  $A$  の出力を  $\ell_{\min}(S, R)$  の多項式時間で計算できるときをいう。

概念クラス  $C$  に対して,  $C^{[n]} = \{c \cap \Upsilon^{[n]} \mid c \in C\}$  と定める。

**定義 20**  $C$  を概念クラスとする. このとき,  $C$  が集合  $S \subseteq \Upsilon^*$  を粉砕するとは, 集合  $\{c \cap S \mid c \in C\}$  が  $S$  の部分集合の全体, すなわち  $2^S$  と等しいときをいう.  $C$  の Vapnik-Chervonenkis 次元, または VC 次元とは,  $C$  によって粉砕される集合  $S$  の濃度の最大値である.  $C$  が多項式次元であるとは, ある多項式  $d(n)$  が存在して,  $C^{[n]} = \{c \cap \Upsilon^{[n]} \mid c \in C\}$  の VC 次元が  $d(n)$  以下であるときをいう.  $C$  の VC 次元が多項式次元でないとき,  $C$  の VC 次元を指数次元という。

**補題 1** (Blumer ら [3], Haussler ら [8], Natarajan [13, 14])  $C$  を概念クラス,  $R$  を多項式時間計算可能な  $C$  の表現とする. このとき,  $C$  が多項式次元かつ  $R$  における  $C$  の多項式時間適合が持てば,  $C$  は  $R$  において多項式時間学習可能である。

**補題 2** (Natarajan [13]) 概念クラス  $C$  が多項式次元であることの必要十分条件は, 多項式  $\text{poly}(n)$  が存在して,  $\log_2 |C^{[n]}| \leq \text{poly}(n)$  ( $n \geq 0$ ) となるときをいう。

### 3 多項式時間学習可能な遺伝的形式グラフ体系

#### 3.1 木幅定数変数独立遺伝的形式グラフ体系

定義 21 グラフ書き換え規則

$$p(g_1, \dots, g_n) \leftarrow q_1(f_1, \dots, f_{\ell_1}), q_2(f_{\ell_1+1}, \dots, f_{\ell_2}), \dots, q_{\ell}(f_{\ell_{\ell-1}+1}, \dots, f_{\ell_{\ell}})$$

が遺伝的であるとは、各項グラフ  $f_j$  ( $1 \leq j \leq \ell_i$ ) がある項グラフ  $g_j$  ( $1 \leq j \leq n$ ) の連結部分項グラフになっているときをいう。

定義 22 グラフ書き換え規則

$$p(g_1, \dots, g_n) \leftarrow q_1(f_1, \dots, f_{\ell_1}), q_2(f_{\ell_1+1}, \dots, f_{\ell_2}), \dots, q_{\ell}(f_{\ell_{\ell-1}+1}, \dots, f_{\ell_{\ell}})$$

が変数遺伝的であるとは、各項グラフ  $f_j$  ( $1 \leq j \leq \ell_i$ ) がある項グラフ  $g_j$  ( $1 \leq j \leq n$ ) に含まれるひとつの変数の星項グラフになっているときをいう。

定義 23  $H\text{-FGS}(m, k, t, r, w, d)$  を、次の条件 (1)–(5) を満たす高々  $m$  節からなる遺伝的形式グラフ体系  $\Gamma$  と単項述語記号  $p$  により定義される形式グラフ体系言語  $GL(\Gamma, p)$  の全体とする。

- (1) ヘッドに現れる変数の個数は高々  $k$  個である。
- (2) ボディに現れるアトム個数は高々  $t$  個である。
- (3) 各述語記号の引数の個数は高々  $r$  個である。
- (4) 各項グラフの超辺のランクは高々  $w$  である。
- (5) 各項グラフの頂点の次数は高々  $d$  個である。

定義 24  $VH\text{-FGS}(m, k, t, r, w, d)$  を、定義 23 の条件 (1)–(5) を満たす高々  $m$  節からなる変数遺伝的形式グラフ体系  $\Gamma$  と単項述語記号  $p$  により定義される形式グラフ体系言語  $GL(\Gamma, p)$  の全体とする。

定義 25  $VH\text{-FGS}^-(m, k, t, r, w, d)$  を、定義 23 の条件 (1)–(5) を満たし、かつ次の条件 (6), (7) を満たす高々  $m$  節からなる変数遺伝的形式グラフ体系



系  $\Gamma$  と単項述語記号  $p$  により定義される形式グラフ体系言語  $GL(\Gamma, p)$  の全体とする。

- (6) 各項グラフの木幅は高々  $w$  である。
- (7) 各項グラフのどの 2 つの超辺もポートを共有しない。

定義 23 の (1)–(5) と定義 25 の (6), (7) を満たす変数遺伝的形式グラフ体系を, 木幅定数変数独立遺伝的形式グラフ体系とよぶ。  $\mathcal{G}(\Sigma, \Lambda)^{[n]}$  を,  $\mathcal{G}(\Sigma, \Lambda)$  の頂点数  $n$  の基礎項グラフの集合とする。また, 各遺伝的形式グラフ体系クラスを次のように定める。

$$\begin{aligned} \text{H-FGS}(m, *, t, *, w, d) &= \bigcup_{k \geq 1} \bigcup_{r \geq 1} \text{H-FGS}(m, k, t, r, w, d), \\ \text{VH-FGS}(m, *, *, r, w, d) &= \bigcup_{k \geq 1} \bigcup_{t \geq 1} \text{VH-FGS}(m, k, t, r, w, d). \end{aligned}$$

図 5 に変数遺伝的形式グラフ体系の例  $\Gamma_{\text{Sierpinski}}$  をあげる。図上の形式グラフ体系は, シェルピンスキ三角形 (Sierpinski triangles) として知られている一連の無向グラフの系列を生成する。  $\Gamma_{\text{Sierpinski}}$  のグラフ書き換え規則より,  $GL(\Gamma_{\text{Sierpinski}}, p) \in \text{VH-FGS}(3, 3, 1, 1, 3, 2)$  である。  $\Gamma_{\text{Sierpinski}}$  中, 2 番目のグラフ書き換え規則のヘッドに現れる項グラフは, 頂点を共有する変数を持つ。このことより,  $\Gamma_{\text{Sierpinski}}$  は変数独立ではない。この共有頂点を 2 つに分離して, 異なる 2 変数が共有頂点を持たないように変更したものが, 図 6 の形式グラフ体系  $\Gamma_{\text{Pseudo-Sierpinski}}$  である。  $\Gamma_{\text{Pseudo-Sierpinski}}$  のグラフ書き換え規則より,  $GL(\Gamma_{\text{Pseudo-Sierpinski}}, p) \in \text{VH-FGS}^-(3, 3, 1, 1, 3, 2)$  が成り立つ。

### 3.2 遺伝的形式グラフ体系の VC 次元

**定理 1**  $\text{H-FGS}(m, *, t, *, w, d)$  は,  $m \geq 10$ ,  $t \geq 1$ ,  $w \geq 2$ ,  $d \geq 2$  のとき, 指数次元である。

**証明.** 基本形式体系に関する指数次元の定理 (Miyano ら [11], Theorem

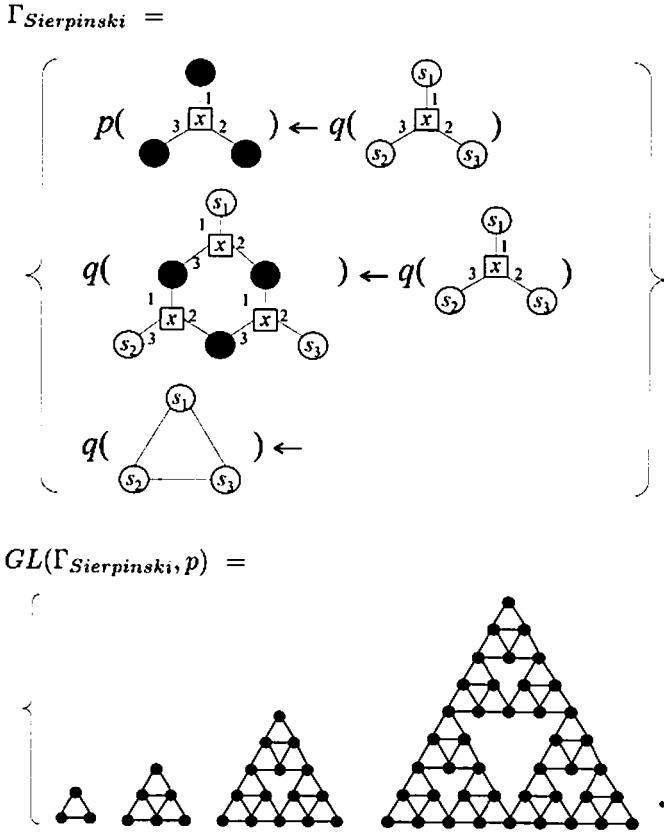
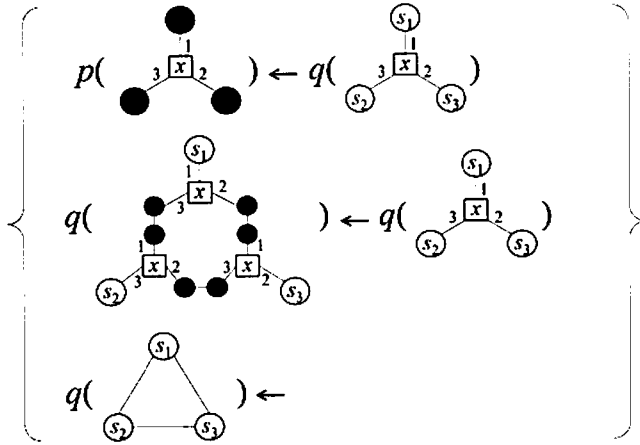


図5 シェルピンスキ三角形を生成する形式グラフ体系  $\Gamma^{Sierpinski}$  とそのグラフ言語  $GL(\Gamma^{Sierpinski}, p)$

4.2) より自明に導かれる。 □

定理2 VH-FGS( $m, *, *, r, w, d$ ) は, 任意の  $m \geq 0, r \geq 0, w \geq 0, d \geq 0$  に対して, 多項式次元である。

$$\Gamma_{Pseudo-Sierpinski} =$$



$$GL(\Gamma_{Pseudo-Sierpinski}, p) =$$

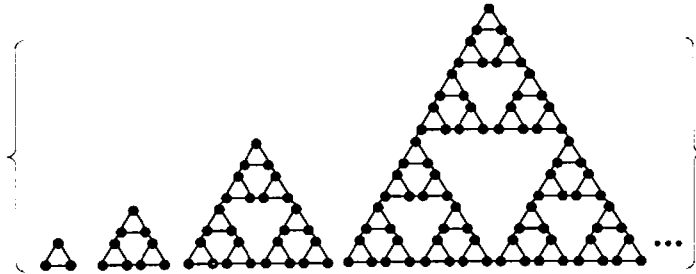


図 6 形式グラフ体系  $\Gamma_{Pseudo-Sierpinski}$  とそのグラフ言語  $GL(\Gamma_{Pseudo-Sierpinski}, p)$

証明. 任意の  $n \geq 0$  に対して,  $VH-FGS(m, *, *, r, w, d)^{[n]}$  を

$$\{L \cap \mathcal{G}(\Sigma, \Lambda)^{[n]} \mid L \in VH-FGS(m, *, *, r, w, d)\}$$

とおく. 以降,  $VH-FGS(m, *, *, r, w, d)^{[n]}$  に含まれる変数遺伝的形式グラフ体系言語の個数を評価する.  $(\Gamma, p)$  を変数遺伝的形式グラフ体系  $\Gamma$  とそ

れに現れる単項述語記号  $p$  のペアとする. グラフ書き換え規則の個数は  $m$  を超えないから, 高々  $r$  の引数を持つ  $m$  個の述語記号だけを考えればよい.  $C = q_0(g_1^0, \dots, g_{r_0}^0) \leftarrow q_1(g_1^1, \dots, g_{r_1}^1), \dots, q_t(g_1^t, \dots, g_{r_t}^t)$  を  $\Gamma$  のグラフ書き換え規則のひとつとする. ある空でない集合  $Y \subseteq \mathcal{G}(\Sigma, \Lambda)^{[n]}$  に関して極小な変数遺伝的形式グラフ体系だけを考えればよい. 従って, 高々頂点数  $n$  の項グラフを含むヘッドから成るグラフ書き換え規則だけを考えなければならぬ.  $n$  頂点の項グラフに対して, 大きさ高々  $w$  のポート順を考慮した超辺の最大個数は,  $\sum_{\ell=1}^w \ell! \binom{n}{\ell} \leq wn^w$  である. ヘッドの述語記号の引数の個数は  $r$  で抑えられるから, 可能な変数の個数は高々  $rwn^w$ , さらに可能なヘッドの数は高々

$$m \cdot |\Sigma|^{nr} \cdot (|\Lambda| + 1)^{\frac{dn}{2}r} \cdot (rwn^w + 1)^{rwn^w}$$

である. 変数遺伝的形式グラフ体系のボディの各項グラフはヘッドのいずれかの項グラフの変数の星項グラフである. ヘッドから得られる星項グラフの個数は高々  $rwn^w$  である. 従って, ボディに現れる可能なアトム個数は高々  $m(rwn^w)^r$  である. ボディのアトムの順序は無視できるので, あるヘッドに対して可能なボディの個数は高々  $2^{m(rwn^w)^r}$  である. 以上より, 可能なグラフ書き換え規則の個数は

$$m \cdot |\Sigma|^{nr} \cdot (|\Lambda| + 1)^{\frac{dn}{2}r} \cdot (rwn^w + 1)^{rwn^w} \cdot 2^{m(rwn^w)^r}$$

で抑えられる. それ故, 変数遺伝的形式グラフ体系の個数は高々

$$(m \cdot |\Sigma|^{nr} \cdot (|\Lambda| + 1)^{\frac{dn}{2}r} \cdot (rwn^w + 1)^{rwn^w} \cdot 2^{m(rwn^w)^r})^m$$

であるから, 次の式が成り立つ.

$$\begin{aligned} & \log_2 |\text{VH-FGS}(m, *, *, r, w, d)^{[n]}| \\ &= m \log_2 m + m r n \log_2 |\Sigma| \\ & \quad + \frac{m r d}{2} n \log_2 (|\Lambda| + 1) + m r w n^w \log_2 (r w n^w + 1) + m^2 (r w n^w)^r \\ &= O(n^{wr} \log n). \end{aligned}$$

従って, 補題 2 より,  $\text{VH-FGS}(m, *, *, r, w, d)$  は多項式次元である.  $\square$

系 1  $\text{VH-FGS}^-(m, *, *, r, w, d)$  は, 任意の  $m \geq 0, r \geq 0, w \geq 0, d \geq 0$  に対して, 多項式次元である.

### 3.3 木幅定数変数独立遺伝的形式グラフ体系の多項式時間適合

補題 3 形式グラフ体系  $\Gamma \in \text{VH-FGS}^-(m, k, t, r, w, d)$  と頂点数  $n$  の基礎項グラフ  $G$  が与えられたとき,  $G \in L(\Gamma, p)$  であるか否かを決定する問題は

$$O(m^3 t r^2 w^4 k d^{2k} n^{2k(w+d)} \cdot \text{ISO}(n)) \text{ 時間}$$

で計算可能である. ここで,  $\text{ISO}(n)$  は頂点数  $n$  の 2 つの基礎項グラフの同型性を判定するために必要な計算時間である.

証明.  $\Gamma \vdash p(G) \leftarrow$  であること, すなわち  $\Gamma$  から  $p(G)$  が証明可能であることの証明を得るために,  $C \in \Gamma$  を使うときはいつも, モーダスポネンスを適用する前に代入  $\theta$  によって基礎グラフ書き換え規則  $C\theta$  を得ることができる. 何故なら,  $p(G)$  は基礎アトムであり,  $\Gamma$  は変数遺伝的であるからである. さらに,  $\Gamma$  が遺伝的であることから, 代入は全ての変数を, 境界頂点の個数が高々  $w$  である  $G$  の連結部分基礎項グラフに対応させるものだけを考慮すればよい. そのような代入により  $\Gamma$  から基礎グラフ書き換え規則の全体  $\mathcal{G}(G)$  を計算する.  $G$  の連結部分基礎項グラフの個数は, 定義 15 の部分グラフの境界表現と定義 25 の (7) より, 高々

$$2wn^w \cdot w \sum_{\ell=0}^d \binom{n}{n-\ell} \leq 2w^2(d+1)n^{w+d}$$

であることがわかる.  $\Gamma$  の各グラフ書き換え規則に対して, 全ての項グラフを  $G$  の部分連結基礎項グラフに対応させる代入の個数は高々  $(2w^2(d+1)n^{w+d})^k$  である. それ故に,  $|\mathcal{G}(G)| \leq m(2w^2(d+1)n^{w+d})^k$  である. グラフ書き換え規則  $C$  とアトム  $A$  に対して, モーダスポネンスを適用することで得られる節は  $O(|C| + |A|)$  時間で計算できる.  $\mathcal{G}(G)$  の各節の長さは高々  $|G|r(1+t)$  である. ここで  $|G|$  はグラフ  $G$  の表現サイズである.

$\mathcal{G}(G)$  にアトム  $A$  が含まれていれば,  $\mathcal{G}(G)$  の節のボディにおける  $A$  の全出現を削除する. このステップは  $O(m \cdot |\mathcal{G}(G)| \cdot \text{ISO}(n) \cdot r^2(1+t)) = O(mtr^2 \cdot |\mathcal{G}(G)| \cdot \text{ISO}(n))$  時間で計算できる. ここで  $\text{ISO}(n)$  は頂点数  $n$  の基礎項グラフと他の基礎項グラフが同型か否かを判定するために必要な計算時間である. このステップをアトム  $p(G)$  が得られるか, もはや新しいアトムが得られなくなるまで繰り返す. これは高々  $|\mathcal{G}(G)|$  ステップで十分だから, 全ての計算は  $O(mtr^2 \cdot |\mathcal{G}(G)|^2 \cdot \text{ISO}(n)) = O(m^3 tr^2 w^{4k} d^{2k} n^{2k(w+d)} \cdot \text{ISO}(n))$  時間で終了する.  $\square$

補題 4 (Bodlaender [4], 名古屋ら [12], 戸田 [18]) 頂点数  $n$ , 木幅高々  $k$  の 2 つのグラフ  $G_1, G_2$  に対して,  $G_1$  と  $G_2$  が同型か否かを判定する問題は  $O(\max\{f(k) \cdot n + k \cdot n^2, n^{k+2}\} + n^{k+4})$  時間で計算可能である. ここで,  $f(k)$  は  $k$  の 2 重指数関数である.

補題 5 形式グラフ体系  $\Gamma \in \text{VH-FGS}^-(m, k, t, r, w, d)$  と頂点数  $n$  の基礎項グラフ  $G$  が与えられたとき,  $G \in L(\Gamma, p)$  であるか否かを決定する問題は  $O(m^3 tr^2 w^{4k} d^{2k} n^{2k(w+d)} \cdot \max\{f(k) \cdot n + k \cdot n^2, n^{k+2}\} + n^{k+4})$  時間で計算可能である. ここで,  $f(k)$  は  $k$  の 2 重指数関数である.

証明. 補題 3 と補題 4 より導かれる.  $\square$

補題 6 形式グラフ体系言語クラス  $\text{VH-FGS}^-(m, k, t, r, w, d)$  は, 任意の非負整数  $m, k, t, r, w, d$  に対して, 多項式時間適合を持つ.

証明.  $S$  を  $\text{VH-FGS}^-(m, k, t, r, w, d)$  の概念  $c$  に対する例の有限集合とする. もし  $S_+$  が空ならば,  $S_-$  の要素でない木幅高々  $w$  の基礎項グラフ  $G \in \Sigma^+$  をひとつ選び,  $\Gamma = \{p(G)\}$  とする. 明らかに  $\Gamma$  は木幅高々  $w$  の項グラフから成る変数遺伝的基本形式体系であり, しかも  $L(\Gamma, p)$  は  $S$  に無矛盾である. それ故, ここで我々は  $S_+$  は空でないと仮定してよい.

$\mathcal{G}(m, k, t, r, w, d, S_+)$  を次の条件 (1)-(3) を満たす全ての形式グラフ体系と述語記号のペア  $(\Gamma, p)$  の全体とする.

1.  $\Gamma$  は, 高々  $m$  個の木幅高々  $w$  の項グラフから成る変数遺伝的グラフ

書き換え規則で、各節のヘッドは変数が高々  $k$  回出現し、ボディは高々  $t$  個の原子を持ち、各述語記号は高々  $r$  個の引数を持つ。

2.  $\Gamma$  の各グラフ書き換え規則の各項グラフ  $g$  に対して、 $g$  への代入  $\theta$  で、 $g\theta$  が  $S_+$  に含まれる基礎項グラフと部分項グラフ同型になるものが存在する。
3. 変数の集合を  $\{x_1, \dots, x_k\}$ 、述語記号の集合を  $\{p, p_1, \dots, p_{m-1}\}$  とする。述語記号  $p$  の引数の個数は 1 に固定する。他の述語記号  $p_i$  ( $i = 1, \dots, m-1$ ) の引数の個数は固定しないでよい。つまり、 $p_i$  ( $i = 1, \dots, m-1$ ) の引数は形式グラフ体系によって異なってよい。

主張 1.  $L(\Gamma, p)$  と  $S$  が無矛盾であるようなペア  $(\Gamma, p) \in \mathcal{G}(m, k, t, r, w, d, S_+)$  が存在する。

主張 1 の証明.  $S$  は  $\text{VH-FGS}^-(m, k, t, r, w, d)$  の概念  $c$  の例の集合であるから、形式グラフ体系  $\Gamma_0$  と述語記号  $p$  のペア  $(\Gamma_0, p)$  で、条件 (1) を満たし、かつ  $L(\Gamma_0, p) = c$  であるものが存在する。ここで、一般性を失わずに、 $\Gamma_0$  は  $S_+$  の正例を生成するために必要なグラフ書き換え規則だけを含んでいるとしてよい。従って、 $\Gamma_0$  は条件 (2) を満たす。さらに  $\Gamma_0$  は高々  $m$  個のグラフ書き換え規則を持つことから、 $\Gamma_0$  には高々  $m$  個の異なる述語記号が存在する。それ故、 $\Gamma_0$  の変数と述語記号の名前を条件 (3) を満たすように変更することができる。(主張 1 の証明終)

主張 2.  $|\mathcal{G}(m, k, t, r, w, d, S_+)|$  は  $|S_+|$  と  $n = \max\{|V(G)| \mid G \in S_+\}$  の多項式で抑えられる。ここで、 $V(G)$  は基礎項グラフ  $G$  の頂点集合を表す。

主張 2 の証明.  $\Pi(k, S_+)$  を次の 2 つの条件を満たす項グラフ  $g$  の集合とする。

- (1)  $g$  は高々  $k$  回の変数出現を持つ項グラフで、変数は  $\{x_1, \dots, x_k\}$  に属す。
- (2) ある代入  $\theta$  が存在して、 $g\theta$  は少なくともひとつの正例  $G \in S_+$  と部分項グラフ同型である。

従って、次の式が成り立つ。

$$\begin{aligned} |\Pi(k, S_+)| &\leq \sum_{G \in S_+} \left( 2w|V(G)|^w \cdot w \sum_{\ell=0}^d \binom{|V(G)|}{|V(G)|-\ell} \right)^{k+1} k^k \\ &\leq \sum_{G \in S_+} (2w^2(d+1)|V(G)|^{w+d})^{k+1} k^k. \end{aligned}$$

この式より、 $|\Pi(k, S_+)| = O(|S_+|(2w^2(d+1)n^{w+d})^{k+1}k^k)$  が成り立つ。可能性のあるヘッドの個数は高々  $m|\Pi(k, S_+)|^r$  である。また、可能性のあるボディのアトム個数は高々  $m(kr)^r$  である。何故なら、ボディの各項グラフはヘッドの項グラフに現れる変数の星項グラフであるからである。従って、次の式が成り立つ。

$$\begin{aligned} |\mathcal{G}(m, k, t, r, w, d, S_+)| &= O((m(|S_+|(2w^2(d+1)n^{w+d})^{k-1}k^k)^r (m(kr)^r)^t)^m) \\ &= O((m^{t+1}|S_+|^r (2w^2(d+1))^{(k+1)r} k^{(k+t)r} r^{rt} m^t n^{(k-1)r(w+d)})^m) \end{aligned}$$

よって、 $m, k, t, r, w, d$  を定数とみなすならば、 $|\mathcal{G}(m, k, t, r, w, d, S_+)|$  は  $|S_+|$  と  $n$  の多項式である。(主張2の証明終)

以上より、次のアルゴリズムは必要な木幅定数遺伝的形式グラフ体系を多項式時間で見つける。 $\mathcal{G}(m, k, t, r, w, d, S_+)$  が含むペア  $(\Gamma, p)$  を列挙する。その都度、全ての  $G \in S_+$  に対して、 $G \in L(\Gamma, p)$  が成り立ち、全ての  $G \in S_-$  に対して、 $G \notin L(\Gamma, p)$  が成り立つかを確認する。この確認は、補題5より多項式時間計算可能である。もし、この確認が全て成り立つペアが見つければ、アルゴリズムはそのペアを仮説として出力する。□

### 3.4 主定理

**定理3** 形式グラフ体系言語クラス  $\text{VH-FGS}^-(m, k, t, r, w, d)$  は多項式時間学習可能である。

証明. 補題1, 系1, 補題6より導かれる。□



## 4 おわりに

本論文では、形式グラフ体系言語クラス  $FGSL$  のサブクラスである木幅定数変数独立遺伝的形式グラフ体系言語クラス  $VH-FGS^-(m, k, t, r, w, d)$  を定義し、そのクラスが多項式時間学習可能であることを証明した。今後の課題としては、まず、木幅定数変数独立遺伝的形式グラフ体系言語クラス  $VH-FGS^-(m, k, t, r, w, d)$  の極小最適教師 (Minimally Adequate Teacher, 略して MAT とよぶ) 学習可能性を明らかにすることがあげられる。MAT 学習とは, Angluin [1] により導入された所属性質問と等価性質問だけをオラクルとして用いて学習する計算論的学習モデルである。

計算論的学習理論の学習モデルとしては、本論文で論じた多項式時間学習モデルの他にも、MAT 学習を拡張した質問学習モデル、正例からの多項式時間帰納推論モデルが知られている。最近, Shibataru ら [17] は、文脈決定文脈自由言語クラスのサブクラスが多項式時間学習可能であることを示した。また, Hara ら [7] は、文脈決定正規形式グラフ体系を定義し、その形式グラフ体系で定義されるグラフ言語が多項式時間 MAT 学習可能であることを示した。これらの結果を踏まえて、今後の興味深い課題として、文脈決定正規形式グラフ体系言語クラスの多項式時間学習可能性の研究、また, Uchida ら [19] が導入したサイズ制限形式グラフ体系に基づく概念クラスの多項式時間学習可能性の研究があげられる。

## 参考文献

- [1] D. Angluin. Learning regular sets from queries and counterexamples. *Information and computation*, Vol.75, No.2, pp.87–106, 1987.
- [2] S. Arikawa, T. Shinohara, and A. Yamamoto. Learning elementary formal systems. *Theoretical Computer Science*, Vol.95, pp.97–113, 1992.
- [3] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth. Learn-

- ability and the Vapnik-Chervonenkis Dimension. *Journal of the ACM*, Vol.36, No.4, pp.929–965, 1989.
- [4] H. L. Bodlaender. Polynomial algorithm for graph isomorphism and chromatic index on partial  $k$ -trees. *Journal of Algorithms*, Vol.11, pp.631–643, 1990.
- [5] D. Chiang, J. Andreas, D. Bauer, K. M. Hermann, B. Jones, and K. Knight. Parsing graphs with hyperedge replacement grammars. *Proceedings of the 51st Meeting of the ACL*. Marie-Catherine de Marneffe, 2013.
- [6] M. Cygan, J. Nederlof, M. Pilipczuk, JMM van Rooij, and J. Wojtaszczyk. Solving connectivity problems parameterized by treewidth in single exponential time. *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science*, pp.150–159, 2011.
- [7] S. Hara and T. Shoudai. Polynomial Time MAT Learning of C-Deterministic Regular Formal Graph Systems. *Proceedings of the 3rd International Conference on Advanced Applied Informatics*, pp.204–211, 2014.
- [8] D. Haussler, M. Kearns, N. Littlestone, and M. K. Warmuth. Equivalence of Models for Polynomial Learnability. *Information and Computation*, Vol.95. No.2, 1991.
- [9] T. Horváth, J. Ramon, and S. Wrobel. Frequent subgraph mining in outerplanar graphs. *Data Mining and Knowledge Discovery*, Vol.21, pp.472–508, 2010.
- [10] J. W. Lloyd. *Foundations of Logic Programming (2nd Extended Ed.)*. Springer-Verlag, New York, Inc., New York, NY, USA, 1987.
- [11] S. Miyano, A. Shinohara, and T. Shinohara. Polynomial-time learning of elementary formal systems. *New Generation Computing*, Vol.18, pp.217–242, 2000.
- [12] 名古屋 孝幸, 谷 聖一, 戸田 誠之助. グラフ同型写像の数え上げ問題に対するアルゴリズム. 信学技報, COMP98-07, 1998.

- [13] B. Natarajan. On learning sets and functions. *Machine Learning*, Vol.4, No.1, pp.67–97, 1989.
- [14] B. Natarajan. *Machine Learning – A Theoretical Approach*. Morgan Kaufmann Publishers, 1991
- [15] 榎原 康文, 小林 聡, 横森 貴. 計算論的学習. 培風館, 2001.
- [16] H. Sakamoto, K. Hirata, and H. Arimura. Learning elementary formal systems with queries. *Theoretical Computer Science*, Vol.298, pp.21–50, 2003.
- [17] C. Shibata and R. Yoshinaka. PAC learning of some subclasses of context-free grammars with basic distributional properties from positive data. *Proceedings of the 24th International Conference on Algorithmic Learning Theory*, Springer, Lecture Notes in Artificial Intelligence, Vol.8139, pp.143–157, 2013.
- [18] 戸田 誠之助. グラフ同型性判定問題. 日本大学文理学部叢書 2, 日本大学文理学部, 富山房, 2001.
- [19] T. Uchida, T. Shoudai, and S. Miyano. Parallel algorithms for refutation tree problem on formal graph systems. *IEICE Transactions on Information and Systems*, Vol.78, No.2, pp.99–112, 1995.
- [20] L. Valiant. A theory of the learnable. *Communications of the ACM*, Vol.27, No.11, pp.1134–1142, 1984.
- [21] H. Yamasaki and T. Shoudai. Mining of Frequent Externally Extensible Outerplanar Graph Patterns. *Proceedings of the 7th International Conference on Machine Learning and Applications*, pp.871–876, 2008.
- [22] H. Yamasaki, Y. Sasaki, T. Shoudai, T. Uchida, and Y. Suzuki. Learning block-preserving graph patterns and its application to data mining. *Machine Learning*, Vo.76, No.1, pp.137–173, 2009.
- [23] X. Yan and J. Han. gSpan graph-based substructure pattern mining. *Proceedings of the 2002 IEEE International Conference on Data Mining*, pp.721–724, 2002.