

# グラフ構造の分布情報を用いた有界な次数と木幅の 文脈自由グラフ言語の多項式時間機械学習

松本 哲志<sup>†</sup> 正代 隆義<sup>‡</sup> 鈴木 祐介<sup>\*</sup>

<sup>†</sup>東海大学理学部情報数理学科 <sup>‡</sup>九州国際大学国際関係学部 <sup>\*</sup>広島市立大学大学院情報科学研究科

## 1. はじめに

形式言語における学習理論の研究では、語句の分布情報を利用する分布学習が大きな成果を挙げている。特に、Clark と Eyraud [2] や Yoshinaka [4] は、分布学習が文脈自由言語を超える複雑な文法形式にも拡張できることを示した。本発表では、分布学習が形式グラフ体系と呼ばれる形式体系によるグラフ言語の学習にも適用できることを示す。形式グラフ体系は、Uchida ら [3] によって導入されたグラフを項とする論理プログラム的一种で、文脈自由グラフ文法と等価な部分クラスを持つ。正規形式グラフ体系は文脈自由グラフ文法として知られている超辺置換グラフ文法 (Hyperedge Replacement Graph Grammar, HRG) と等価な形式グラフ体系である[3]。本発表では、正規形式グラフ体系から得られるグラフ言語のクラスは、それが有限な文脈で表現可能で、有界な次数と木幅のグラフ言語を生成するとき、所属性質問を用いて多項式更新時間で正例から極限同定可能であることを示す。

## 2. 項グラフパターン

$\Sigma$  と  $A$  を有限アルファベット、 $X$  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。 $X$  の記号  $x$  に非負整数  $rank(x)$  を仮定する。また、 $\Sigma, A, X$  は、 $\Sigma \cap X = \emptyset$  および  $A \cap X = \emptyset$  が成り立つとする。集合または順序リスト  $S$  に対して、 $|S|$  を  $S$  に含まれる要素の個数とする。

**定義 1 ([3]):** アルファベット  $\langle \Sigma, A, X \rangle$  上の項グラフパターン、または項グラフとは、次の4つの条件を満たす7つ組  $g = (V, E, \partial, \varphi, H, \lambda, ports)$  である。

1.  $(V, E)$  は有向グラフまたは単純無向グラフである。 $\partial$  は頂点ラベル付け関数  $\partial: V \rightarrow \Sigma$  であり、 $\varphi$  は辺ラベル付け関数  $\varphi: E \rightarrow A$  である。
2.  $H$  は空でない  $V$  上の超辺の集合である。
3.  $\lambda$  は超辺からの変数への関数  $\lambda: H \rightarrow X$  で、 $rank(\lambda(h)) = |h|$  を満たす関数である。

4.  $ports$  は関数  $ports: H \rightarrow V^*$  で、各超辺  $h \in H$  に対して、 $ports(h)$  は  $rank(\lambda(h))$  個の異なる頂点からなる順序付きリストである。 $ports(h)$  中の頂点を超辺  $h$  のポートとよぶ。

項グラフの例を図1に与える。この例以降、変数を囲む四角で超辺を表す。超辺のメンバーを四角と線で結ぶ。数字は  $ports$  による頂点順を表す。

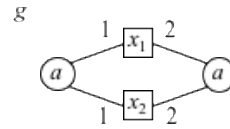


図1 項グラフの例

$H$  が空である項グラフを基礎項グラフとよぶ。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  $[f, \sigma]$  を項グラフ断片とよぶ。特に、 $f$  が基礎項グラフのとき、基礎項グラフ断片とよぶ。

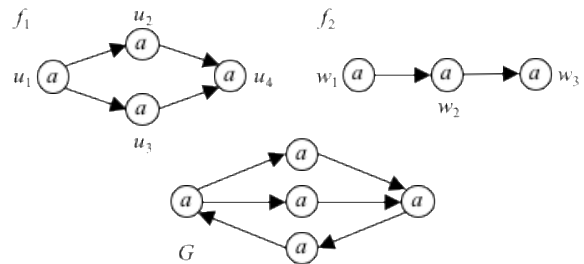


図2 基礎項グラフ

$x$  を  $X$  の変数、 $[f, \sigma]$  を項グラフ断片とする。ただし、 $rank(x) = |\sigma|$ 、 $f = (V_f, E_f, \partial_f, \varphi_f, H_f, \lambda_f, ports_f)$  とする。このとき、形式  $x := [f, \sigma]$  を  $x$  の束縛とよぶ。 $g = (V, E, \partial, \varphi, H, \lambda, ports)$  を項グラフとする。変数  $x$  の束縛を適用して得られる項グラフ  $g\{x := [f, \sigma]\}$  を次のように定める。 $f'$  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。 $g$  の  $x$  を変数ラベルとする超辺  $h = (u_1, \dots, u_k) \in H$  に対して、超辺  $h$  を  $H$  から削除し、 $h$  のポート  $u_1, \dots, u_k$  を  $f'$  の頂点  $v_1', \dots, v_k'$  と、この順で同一視して  $f'$  を  $g$  に貼付ける。多重辺が出来るならば、対応する  $f'$  の辺を削除する。また、 $u_i$  の頂点ラベルを  $\partial(u_i)$  とする。代入とは束縛の有限集合である。項グラフ  $g$  に対して、代入  $\theta$  の束縛を全て適用して得られるグラフを  $g\theta$  と書く。図1の  $g$  と図2の  $f_1, f_2, G$  に対して、 $G$  は  $g$  に  $\{x_1 := [f_1, (u_1, u_4)], x_2 := [f_2, (w_3, w_1)]\}$  を適用して得る。

Distributional Learning of Regular Formal Graph System of Bounded Degree and Bounded Treewidth

<sup>†</sup> Satoshi Matsumoto, Faculty of Science, Tokai University, Japan

<sup>‡</sup> Takayoshi Shoudai, Faculty of International Studies, Kyushu International University, Japan

<sup>\*</sup> Yusuke Suzuki, Graduate School of Information Sciences, Hiroshima City University, Japan

### 3. 形式グラフ体系

$g_1, \dots, g_n$  を  $\langle \Sigma, A, X \rangle$  上の項グラフとする.  $n$  引数述語記号  $p$  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**(Graph Rewriting Rule, GRR)とは,  $A \leftarrow B_1, \dots, B_m$  という形の確定節である. **形式グラフ体系**(Formal Graph System, FGS)とは GRR の有限集合  $\Gamma$  である.  $\theta$  を代入,  $p(g_1, \dots, g_n)$  をアトムとする.  $p(g_1, \dots, g_n)\theta$  を  $p(g_1\theta, \dots, g_n\theta)$  と定める.  $\text{GRR } A \leftarrow B_1, \dots, B_m$  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.  $\Gamma$  を FGS とする.  $\delta$  を述語記号の集合  $\Pi$  から  $\Sigma^*$  への関数とする.  $\delta$  は  $p$  に対して,  $\Sigma$  の互いに異なる記号列  $a_1, \dots, a_k$  ( $k \geq 0$ ) を定める関数 (ポインタと呼ぶ) である.  $\delta(p) = (a_1, \dots, a_k)$  と記述する.  $\delta(\Pi) = \{a \in \Sigma \mid \exists p \in \Pi \text{ s.t. } a \in \delta(p)\}$  とする.

**定義 2 ([3]):** FGS  $\Gamma$  と述語記号のポインタ  $\delta$  が次の条件を満たすとき,  $\Gamma$  と  $\delta$  の組  $(\Gamma, \delta)$  を **正規形式グラフ体系**(Regular FGS, RFGS) と呼ぶ.

- $\Gamma$  の全ての GRR の述語記号は 1 引数で,  $q_0(g_0) \leftarrow q_1(g_1), \dots, q_m(g_m)$  と書ける.
- 各  $g_i$  ( $i=1, \dots, m$ ) は星形項グラフ (全頂点がメンバーである一つの超辺しかない) である.
- 各  $g_i$  ( $i=1, \dots, m$ ) にただ一つ存在する超辺のポートの頂点ラベルの列は  $\delta(q_i)$  と一致する.
- $g_0$  に  $|\delta(q_0)|$  個の頂点があつて,  $\delta(q_0)$  の頂点ラベルが貼られている. その他の頂点には,  $\delta(\Pi)$  以外の頂点ラベルが貼られている.
- $g_0$  の変数の集合は  $g_i$  ( $i \geq 1$ ) の変数の和集合に等しい.  $g_i$  と  $g_j$  ( $i, j \geq 1$ ) の変数は互いに異なる.
- $g_0$  の変数には互いに異なる.

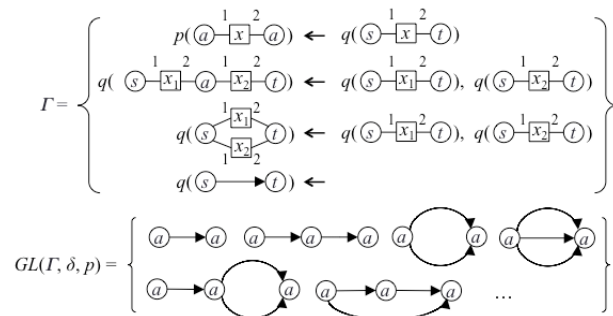


図 3 RFGS とその言語(TTSP グラフ全体)の例

$f_0$  を 1 頂点または 1 辺と 2 頂点からなる基礎項グラフとする.  $f_1$  を 2 超辺だけの項グラフ,  $f_2, f_3$  を星形項グラフとし,  $p_0, p_1, p_2, p_3$  を 1 引数述語記号とする. FGS  $\Gamma$  の GRR が  $p_0(f_0) \leftarrow$  または  $p_1(f_1) \leftarrow p_2(f_2), p_3(f_3)$  のどちらかの形式であるとき,  $\Gamma$  を **チョムスキー標準形** と呼ぶ.

$(\Gamma, \delta)$  を RFGS,  $p$  を 1 引数述語記号とする. **正規形式グラフ体系言語**(RFGS 言語)は, 基礎項グラフ  $g$  の集合で,  $p(g)$  が  $\Gamma$  の GRR と代入, モーダ

スポネンスを使って証明可能な基礎項グラフ  $g$  の全体である. RFGS  $(\Gamma, \delta)$  と 1 引数述語記号  $p$  による RFGS 言語を  $GL(\Gamma, \delta, p)$  と記す. 図 3 に RFGS の例をあげる.

### 4. グラフ構造の分布情報による学習

$x \in X$  をどこにも使われていない新しい変数とする.  $\sigma = (v_1, \dots, v_k)$  を頂点リスト,  $[g, \sigma]$  を基礎項グラフ断片とする.  $g(\sigma) = (V_g, E_g, \partial_g, \varphi_g, \{h\}, \lambda_g, \text{ports}_g)$  と定める.  $h = \{v_1, \dots, v_k\}$ ,  $\lambda_g(h) = x$ ,  $\text{ports}_g(h) = \sigma$  とする.  $g$  は孤立頂点を持たないと仮定する. 基礎項グラフ断片  $[\alpha, \sigma_\alpha]$  と  $[\beta, \sigma_\beta]$  の演算  $\bullet$  を, もし  $|\sigma_\alpha| = |\sigma_\beta|$  ならば,  $[\alpha, \sigma_\alpha] \bullet [\beta, \sigma_\beta] = \alpha(\sigma_\alpha) \{x := [\beta, \sigma_\beta]\}$  と定める. もし  $|\sigma_\alpha| \neq |\sigma_\beta|$  でなければ定義しない.

$(\Gamma, \delta)$  を RFGS,  $f$  を項グラフ,  $q$  を述語記号とする. 全ての  $i$  ( $1 \leq i \leq |\delta(q)|$ ) に対して,  $\exists v_i \in V_f$  s.t.  $\partial_f(v_i) = \delta(q)[i]$  ならば,  $\partial_f^{-1}(\delta(q)) = (v_1, \dots, v_{|\delta(q)|})$  とし, そうでなければ,  $\partial_f^{-1}(\delta(q)) = ()$  とおく.

**定義 3:** RFGS  $(\Gamma, \delta)$ ,  $p, q$  を述語記号,  $[g, \sigma_g]$  を基礎項グラフ断片とする. このとき,

$C(\Gamma, \delta, p, q, g) = \{f \mid [g, \sigma_g] \bullet [f, \partial_f^{-1}(\delta(q))] \in GL(\Gamma, \delta, p)\}$  とおく. RFGS  $(\Gamma, \delta)$  と述語記号  $p$  が **1-有限文脈性** (1-Finite Context Property, 1-FCP) を持つとは, 全ての  $q \in \Pi$  に対して, 基礎項グラフ断片  $[g, \sigma_g]$  が存在して,  $C(\Gamma, \delta, p, q, g) = GL(\Gamma, \delta, q)$  が成り立つときをいう.

**定義 4:**  $w$  および  $d$  を正整数定数とする. **1-FCP-RFGSL( $w, d$ )** を次の条件を満たす RFGS 言語  $L$  の全体とする.

- 1-有限文脈性を持つ RFGS  $(\Gamma, \delta)$  と 1 引数述語記号  $p$  が存在して,  $L = GL(\Gamma, \delta, p)$  となる.
- $\Gamma$  はチョムスキー標準形である.
- $\Gamma$  中の項グラフの木幅[1]は高々  $w$  である.
- $L$  に属するグラフの最大次数は高々  $d$  である.

**定理 1:**  $w$  と  $d$  を正整数定数とする. グラフ言語のクラス **1-FCP-RFGSL( $w, d$ )** は所属性質問を用いて多項式更新時間で正例から極限同定可能である.

### 参考文献

- D. Chiang, et al., "Parsing graphs with hyperedge replacement grammars", Proc. ACL 2013, pages 924-932, 2013.
- A. Clark and R. Eyraud, "Polynomial identification in the limit of substitutable context-free languages", JMLR 8, pp. 1725-1745, 2007.
- T. Uchida, T. Shoudai, and S. Miyano, "Parallel algorithms for refutation tree problem on formal graph systems", Trans. Inf. & Syst. E78-D(2), pp. 99-112, 1995.
- R. Yoshinaka, "Integration of the dual approaches in the distributional learning of context-free grammars", Proc. LATA2012, LNCS 7183, Springer, pages. 538-550, 2012.