

木オートマトンを用いた化学グラフのスクリーニング手法

藤芳 明生

茨城大学工学部情報工学科

fujiyosi@mx.ibaraki.ac.jp

概要

本稿では、木オートマトンによる化学グラフの所属問題を考える。化学グラフが木オートマトンによって受理されることを、その化学グラフが木オートマトンによって受理される全域木を持つことと定義する。この所属問題が NP 完全であることと、Tree-Width が高々2 であるような化学グラフに対する線形時間の所属判定アルゴリズムを提案する。

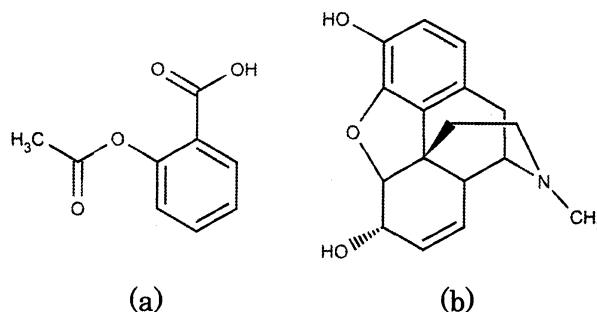


図 1: (a) アセチルサリチル酸, (b) モルヒネ.

1 はじめに

化学グラフ (分子グラフ) とは、化学物質の分子構造 (構造式) を表現したグラフのことである。各頂点は原子の種類でラベル付けされており、各辺は原子間の結合の種類によってラベル付けされている。更に、立体異性体を区別して表現するために、結合の位置関係を考慮しなければならない場合もある。図 1 にアセチルサリチル酸とモルヒネの構造式を示す。要するに、このような構造式をグラフとして見なしたものが、化学グラフであるといえる。

アルゴリズムの視点から化学グラフを考えると、それらは非常に扱いやすいグラフであるといえる [1]。まず、化学グラフの最大次数は定数で制限できることが挙げられる。よく知られているように、炭素原子が共有結合できる数は 4 つまでであり、水素原子は 1 つまで、酸素原子は 2 つまでとなっている。配位結合も考慮するとしても、化学グラフの次数は 6 以下であると制限してよい。最大次数が定数で制限されたグラフに対しては、グラフ同型性判定 [2] 及び

Canonical Labelling [3] が多項式時間でできることが知られている。次に、ほとんどの化学グラフは平面的であるということが挙げられる。ゼオライトの結晶のように、無機化合物には平面的ではないものも存在するが、それらは例外的であると言ってよい。平面的のグラフに対しては、グラフ同型性判定が線形時間 [4] でできることが知られている。更に、医薬品及び医薬品候補化合物に限定すれば、それらの化学グラフの Tree-Width は十分に小さいということが挙げられる。表 1 に、ChEMBL[5] に登録されている 635,933 個の医薬品及び医薬品候補化合物の Tree-Width を示す。ほとんどの化合物の Tree-Width は 3 以下となっていることが確認された。Tree-Width が定数で制限されたグラフに対しては、グラフ同型性判定が多項式時間 [6] でできることが知られているだけでなく、Tree-Width と最大次数が共に定数で制限されたグラフに対しては、部分グラフ同型性判定も多項式時間 [7] でできることが知られている。部分グラフ同型性判定問題は、よく知られているよう

表 1: ChEMBL 中の化合物の Tree-Width.

Tree-Width	数	割合
1	11,753	(1.85%)
2	603,806	(94.95%)
3	20,031	(3.15%)
≥ 4	343	(0.05%)
Total	635,933	(100.00%)

に, 一般のグラフに対しては NP 完全問題である.

本稿は, 化学グラフの所属性判定に木オートマトン [8] を利用することを提案する. 議論が複雑になりすぎるため, 辺のラベル (結合の種類) は無視することとし, 化学グラフを頂点ラベル付グラフと見なすことにする. 頂点ラベル付グラフを木オートマトンが受理することを, そのグラフにある根付き全域木が存在し, 木オートマトンがその根付き全域木を受理することと定義する. まず, 木オートマトンによる頂点ラベル付グラフの所属問題が NP 完全であることを示す. 次に, Tree-Width が高々 2 であるような化学グラフに対しては線形時間の所属性判定アルゴリズムが存在することを示す.

2 諸定義

グラフとは, 順序対 $G = (V, E)$ のことである. ここで, V は頂点の有限集合, 異なる頂点の組を辺と呼び, E は辺の有限集合である. $u, v \in V$ かつ $e = \{u, v\} \in E$ であるとき, u と v は隣接するという. 木とは, 閉路を持たないグラフのことである. $T = (V, E)$ を木, $r \in V$ とする. 根付き木とは, 順序対 $T' = (V, E, r)$ のことである.

$G = (V, E)$ をグラフとする. Σ を頂点ラベルの有限集合とし, 関数 $\sigma: V \rightarrow \Sigma$ を, 頂点ラベル付け関数と呼ぶ. σ が V 上で定義されているとき, G は頂点ラベル付きであると言う. 本稿では, すべてのグラフは頂点ラベル付きであるとする.

$G = (V_1, E_1)$ をグラフ, σ_1 をその頂点ラベル付け

関数とする. $T = (V_2, E_2)$ を木, σ_2 をその頂点ラベル付け関数とする. T が G の全域木であるとは, $V_1 = V_2$, $\sigma_1 = \sigma_2$, かつ, $E_2 \subseteq E_1$ となることである. グラフに全域木が存在するためには, 連結でなくてはならないので, 本稿では, 連結グラフだけを考える.

$T = (V, E, r)$ を根付き木, $\sigma: V \rightarrow \Sigma$ をその頂点ラベル付け関数とする. T に対応する順序木とは, Σ の元と括弧及びコンマを用いて, 次のように再帰的に定義される文字列 w である.

- $V = \{r\}$ かつ $E = \emptyset$ のとき, $w = \sigma(r)$ は T に対応する順序木である.
- $n \geq 1$, $T_1 = (V_1, E_1, r_1), \dots, T_n = (V_n, E_n, r_n)$, $V = V_1 \cup \dots \cup V_n \cup \{r\}$, $r \notin V_1 \cup \dots \cup V_n$, すべての $1 \leq i, j \leq n$ の組に対し $V_i \cap V_j = \emptyset$, $E = E_1 \cup \dots \cup E_n \cup \{\{r, r_1\}, \dots, \{r, r_n\}\}$ かつ w_1, \dots, w_n が T_1, \dots, T_n に対応する順序木するとき, $w = \sigma(r)(w_1, \dots, w_n)$ は T に対応する順序木である.

w_1, \dots, w_n の並び方は任意であるため, T に対応する順序木はその組み合わせの数だけ存在する.

$\mathcal{X} = \{x_1, x_2, \dots\}$ を変数の集合とする.

3 木オートマトン

本章では, 頂点ラベル付グラフを受理する木オートマトンを紹介する. ここで紹介する木オートマトンの定義は, 通常の木オートマトン [8] と全く同じである. 本稿では, トップダウン型の木オートマトンだけを定義するが, 同等の認識能力を持ったボトムアップ型の木オートマトンを定義することもできる.

定義 1 (非決定性トップダウン) 木オートマトンとは, 四つ組 $\mathcal{A} = (Q, \Sigma, q_0, R)$ のことである. ここで, Q は状態の有限集合, $q_0 \in Q$ は初期状態, R は次の形の規則の有限集合である.

$$q(f(x_1, \dots, x_n)) \rightarrow f(q_1(x_1), \dots, q_n(x_n))$$

ここで, $n \geq 0$, $f \in \Sigma$, $q, q_1, \dots, q_n \in Q$, $x_1, \dots, x_n \in \mathcal{X}$ である. $n = 0$ のときは, $q(f()) \rightarrow f()$ の代わりに, $q(f) \rightarrow f$ と書く.

D を頂点ラベル付グラフ, Σ をその頂点ラベルの集合, A を Σ 上の木オートマトンとする. D が A によって受理されるとは, D の根付き全域木 T 及び T に対応する順序木 w が存在し, w が A によって受理されることである.

4 頂点ラベル付グラフの所属問題の NP 完全性

文献 [9] で, 木オートマトンによる無閉路有向グラフ (DAG) の所属問題が NP 完全であることが証明されている. 本稿では, 文献 [9] の証明手順を頂点ラベル付グラフの所属問題の結果に適応し, 次の結果を得る.

補題 1 Σ を頂点ラベルの集合とする. $|\Sigma| \geq 3$ であるならば, 木オートマトンによる頂点ラベル付グラフの所属問題は NP 完全である.

(証明) 充足可能性問題 (SAT) をこの問題に還元できることを示す.

Σ が異なる 3 つのラベル a, b, c を含むと仮定し, 次のような木オートマトン $A = (Q, \Sigma, q_0, R)$ を用意する.

$$\begin{aligned} Q &= \{q_0, q_1, q_2, q_3\}, \\ R &= \{q_0(a(x_1, x_2)) \rightarrow a(q_1(x_1), q_3(x_2)), \\ &\quad q_0(b(x_1, x_2)) \rightarrow b(q_1(x_1), q_3(x_2)), \\ &\quad q_1(b(x_1, x_2)) \rightarrow b(q_0(x_1), q_2(x_2)), \\ &\quad q_1(b(x_1)) \rightarrow b(q_2(x_1)), \\ &\quad q_2(b(x_1, x_2)) \rightarrow b(q_2(x_1), q_2(x_2)), \\ &\quad q_2(b(x_1)) \rightarrow b(q_2(x_1)), q_2(c) \rightarrow c, \\ &\quad q_3(b(x_1)) \rightarrow b(q_3(x_1)), q_3(c) \rightarrow c\}. \end{aligned}$$

\mathcal{F} を与えられた乗法標準形 (CNF) の論理式とする. $\mathcal{C} = \{c_1, \dots, c_m\}$ を \mathcal{F} を構成するクローズの集

合, $\mathcal{V} = \{v_1, \dots, v_n\}$ を \mathcal{F} に現れる変数の集合とする. \mathcal{F} から頂点ラベル付きグラフ $G = (V, E)$ を次のように構成する.

$$\begin{aligned} V &= \{v_1, \dots, v_n\} \cup \{\bar{v}_1, \dots, \bar{v}_n\} \cup \{c_1, \dots, c_m\} \\ &\quad \cup \{s_{v_1}, \dots, s_{v_n}\} \cup \{t_{v_1}, \dots, t_{v_n}, t_{\bar{v}_1}, \dots, t_{\bar{v}_n}\} \\ &\quad \cup \{u_{v_1}, \dots, u_{v_n}, u_{\bar{v}_1}, \dots, u_{\bar{v}_n}\} \\ &\quad \cup \{w_{[v_i, c_j]}, w_{[\bar{v}_i, c_j]} \mid 1 \leq i \leq n, 1 \leq j \leq m\}, \end{aligned}$$

$$\begin{aligned} E &= \{\{s_{v_i}, v_i\}, \{s_{v_i}, \bar{v}_i\} \mid 1 \leq i \leq n\} \\ &\quad \cup \{\{v_i, s_{v_{i+1}}\}, \{\bar{v}_i, s_{v_{i+1}}\} \mid 1 \leq i \leq n-1\} \\ &\quad \cup \{\{v_i, u_{v_i}\}, \{\bar{v}_i, u_{\bar{v}_i}\} \mid 1 \leq i \leq n\} \\ &\quad \cup \{\{u_{v_i}, w_{[v_i, c_1]}\}, \{u_{\bar{v}_i}, w_{[\bar{v}_i, c_1]}\} \mid 1 \leq i \leq n\} \\ &\quad \cup \{\{w_{[v_i, c_j]}, w_{[v_i, c_{j+1}]}\}, \{w_{[\bar{v}_i, c_j]}, w_{[\bar{v}_i, c_{j+1}]}\} \mid \\ &\quad \quad 1 \leq i \leq n, 1 \leq j \leq m-1\} \\ &\quad \cup \{\{w_{[v_i, c_m]}, t_{v_i}\}, \{w_{[\bar{v}_i, c_m]}, t_{\bar{v}_i}\} \mid 1 \leq i \leq n\} \\ &\quad \cup \{\{w_{[v_i, c_j]}, c_j\} \mid \\ &\quad \quad 1 \leq i \leq n, 1 \leq j \leq m, v_i \text{ が } c_j \text{ に存在する}\} \\ &\quad \cup \{\{w_{[\bar{v}_i, c_j]}, c_j\} \mid \\ &\quad \quad 1 \leq i \leq n, 1 \leq j \leq m, \bar{v}_i \text{ が } c_j \text{ に存在する}\}. \end{aligned}$$

頂点ラベル付け関数は, $\sigma(s_{v_1}) = a$, $v \in \{c_1, \dots, c_m\} \cup \{t_{v_1}, \dots, t_{v_n}, t_{\bar{v}_1}, \dots, t_{\bar{v}_n}\}$ に対し, $\sigma(v) = c$, それ以外の $v \in V$ に対し, $\sigma(v) = b$ と定義する.

例として, 論理式 $(v_1 \vee \bar{v}_2 \vee v_3) \wedge (v_1 \vee v_3 \vee \bar{v}_4) \wedge (\bar{v}_2 \vee \bar{v}_3 \vee v_4)$ を考える. 図 2 に, この論理式に対応する頂点ラベル付きグラフを示す. a のラベルが付いた頂点は斜線, b のラベルが付いた頂点は白丸, c のラベルが付いた頂点は黒丸で表してある.

明らかに, G が A に受理される必要十分条件は, 論理式 \mathcal{F} に充足解が存在することである. よって, 木オートマトンによる頂点ラベル付きグラフの全域木の所属問題は NP 困難である.

一方, 明らかに, 与えられた頂点ラベル付きグラフ G に対し, 非決定的多項式時間で G の全域木 T を構成し, T が A に受理されるかどうか確認することができる. よって, 木オートマトンによる頂点ラ

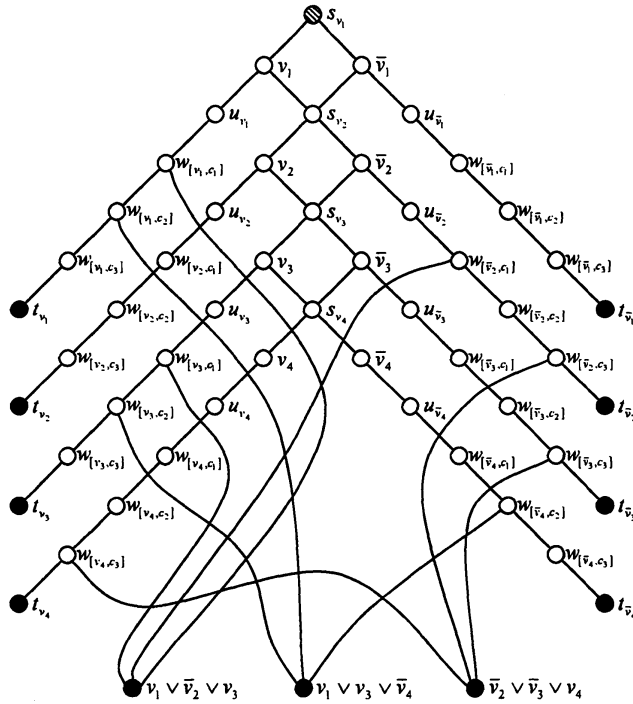


図 2: 論理式 $(v_1 \vee \bar{v}_2 \vee v_3) \wedge (v_1 \vee v_3 \vee \bar{v}_4) \wedge (\bar{v}_2 \vee \bar{v}_3 \vee v_4)$ に対応するグラフ.

ベル付きグラフの全域木の所属問題はクラス NP に属する.

ゆえに、 $|\Sigma| \geq 3$ であるならば、木オートマトンによる頂点ラベル付きグラフの全域木の所属問題は NP 完全である。□

文献 [9] では、任意のラベルの集合に対し、木オートマトンによる無閉路有向グラフ (DAG) の所属問題が NP 完全であることが証明されている。しかしながら、補題 1 は、ラベルが 3 種類以上ある場合の結果である。そこで、補題 1 の結果を拡張し、次の結果を得る。

定理 1 任意の頂点ラベルの集合に対し、木オートマトンによる頂点ラベル付きグラフの所属問題は NP 完全である。

(証明) 少なくとも頂点ラベルは 1 種類はあるはずなので、 $f \in \Sigma$ と仮定する。補題 1 で構成したグラフ G に対し、次のような操作を行う。

- a のラベルの付いた頂点は、そのままにする。
- それぞれの b のラベルの付いた頂点に対し、新しい頂点を 2 つ用意し、それらを繋げる。
- それぞれの c のラベルの付いた頂点に対し、新しい頂点を 5 つ用意し、それらを繋げる。
- 最後に、すべての頂点のラベルを f にする。

新しくできたグラフを G' とする。 a のラベルの付いた頂点は、 G' でも次数が 2 のままであり、 b のラベルの付いた頂点は、 G' では次数が 4 以上 6 以下になり、 c のラベルの付いた頂点は、 G' では次数が 8 以上になる。よって、ラベルの種類がすべて同じであっても、木オートマトンは区別して動作することができる。

次のような木オートマトン $A' = (Q', \Sigma, q_0, R')$ を用意する。

$$\begin{aligned}
 Q' &= \{q_0, q_1, q_2, q_3, q_4\}, \\
 R' &= \{q_0(f(x_1, x_2)) \rightarrow f(q_1(x_1), q_3(x_2)), \\
 &\quad q_0(f(x_1, x_2, x_3, x_4)) \\
 &\quad \rightarrow f(q_1(x_1), q_3(x_2), q_4(x_3), q_4(x_4)), \\
 &\quad q_1(f(x_1, x_2, x_3, x_4)) \\
 &\quad \rightarrow f(q_0(x_1), q_2(x_2), q_4(x_3), q_4(x_4)), \\
 &\quad q_1(f(x_1, x_2, x_3)) \\
 &\quad \rightarrow f(q_2(x_1), q_4(x_2), q_4(x_3)), \\
 &\quad q_2(f(x_1, x_2, x_3, x_4)) \\
 &\quad \rightarrow f(q_2(x_1), q_2(x_2), q_4(x_3), q_4(x_4)), \\
 &\quad q_2(f(x_1, x_2, x_3)) \\
 &\quad \rightarrow f(q_2(x_1), q_4(x_2), q_4(x_3)), \\
 &\quad q_3(f(x_1, x_2, x_3)) \\
 &\quad \rightarrow f(q_3(x_1), q_4(x_2), q_4(x_3)), \\
 &\quad q_2(f(x_1, x_2, x_3, x_4, x_5)) \\
 &\rightarrow f(q_4(x_1), q_4(x_2), q_4(x_3), q_4(x_4), q_4(x_5)), \\
 &\quad q_3(f(x_1, x_2, x_3, x_4, x_5)) \\
 &\rightarrow f(q_4(x_1), q_4(x_2), q_4(x_3), q_4(x_4), q_4(x_5)), \\
 &\quad q_4(f) \rightarrow f\}.
 \end{aligned}$$

明らかに、 G' が A' に受理されることは、補題1の G が A に受理されることと同値である。□

5 Tree-Width が高々2 のグラフ に対する線形時間の所属性判定 アルゴリズム

本章では、Tree-Width が高々2 のグラフに対する線形時間の所属性判定アルゴリズムを紹介する。

任意の Tree-Width が高々2 のグラフは、次に示す変形規則によって単一頂点のグラフに変形できることが知られている [10] (図3 も見よ)。

1. グラフに頂点 v_2 が存在し、正確に1つの辺 $\{v_1, v_2\}$ に繋がっているとき、頂点 v_2 と辺 $\{v_1, v_2\}$ を取り除く。
2. グラフに頂点 v_2 が存在し、正確に2つの辺 $\{v_1, v_2\}, \{v_2, v_3\}$ に繋がっているとき、頂点 v_2 と辺 $\{v_1, v_2\}, \{v_2, v_3\}$ を取り除き、新しい辺 $\{v_1, v_3\}$ を加える。
3. 規則2によって、頂点 v_1, v_2 の間に多重辺ができてしまったら、多重辺を取り除き、新しい辺 $\{v_1, v_3\}$ を加える。

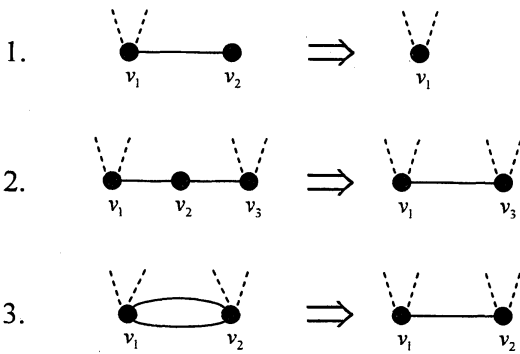


図3: 変形規則.

本稿で紹介する所属性判定アルゴリズムは、変形規則にグラフが変形される毎に、各頂点及び辺に与えた値を書き換えていくことで動作する。

$A = (Q, \Sigma, q_0, R)$ を与えられた木オートマトンとする。 \mathcal{X}_A を R に現れる変数の集合とする。次の形の規則 $r \in R$ に対し、

$$r: q(f(x_1, \dots, x_n)) \rightarrow f(q_1(x_1), \dots, q_n(x_n))$$

$\text{state}(r)$ で状態 q を表し、 $\text{sym}(r)$ で記号 f を表し、 $\text{var}(r)$ で変数の集合 $\{x_1, \dots, x_n\}$ を表すものとする。

所属性判定アルゴリズム

入力: 頂点ラベル付グラフ $G = (V, E)$ とその頂点ラベル付け関数 σ

出力: *accept* または *reject*

- 1: 頂点集合 V に任意の全順序を与える
- 2: $E' := \{(v_1, v_2) \mid \{v_1, v_2\} \in E \text{ かつ } v_1 < v_2\}$
- 3: **for all** $e: (v_1, v_2) \in E'$ **do**
- 4: $A[e] := \emptyset$
- 5: $B[e] := \emptyset$
- 6: $C[e] := \emptyset$
- 7: $D[e] := \emptyset$
- 8: **for all** $r: q(f(x_1, \dots, x_n)) \rightarrow f(q_1(x_1), \dots, q_n(x_n)) \in R$ **do**
- 9: **if** $n \geq 1$ かつ $\sigma(v_1) = f$ **then**
- 10: $A[e] := A[e] \cup \{(r, \{x_i\}, r', \emptyset) \mid 1 \leq i \leq n, r' \in R, \text{sym}(r') = \sigma(v_2) \text{ かつ } \text{state}(r') = q_i\}$
- 11: **end if**
- 12: **if** $n \geq 1$ かつ $\sigma(v_2) = f$ **then**
- 13: $B[e] := B[e] \cup \{(r', \emptyset, r, \{x_i\}) \mid 1 \leq i \leq n, r' \in R, \text{sym}(r') = \sigma(v_1) \text{ かつ } \text{state}(r') = q_i\}$
- 14: **end if**
- 15: **end for**
- 16: **for all** $(r, r') \in R \times R$ **do**
- 17: **if** $\text{sym}(r) = \sigma(v_1)$ かつ $\text{sym}(r') = \sigma(v_2)$ **then**
- 18: $C[e] := C[e] \cup \{(r, \emptyset, r', \emptyset)\}$
- 19: **end if**
- 20: **end for**
- 21: **end for**
- 22: **for all** $v \in V$ **do**

23: $A[v] := \emptyset$
 24: $B[v] := \emptyset$
 25: **for all** $r \in R$ **do**
 26: **if** $\text{sym}(r) = \sigma(v)$ **then**
 27: $A[v] := A[v] \cup \{(r, \emptyset)\}$
 28: **end if**
 29: **if** $\text{state}(r) = q_0$ **かつ** $\text{sym}(r) = \sigma(v)$ **then**
 30: $B[v] := B[v] \cup \{(r, \emptyset)\}$
 31: **end if**
 32: **end for**
 33: **end for**
 34: **while** $|V| > 1$ **do**
 35: **if** $v_2 \in V$ が存在し、正確に1つの辺 $\{v_1, v_2\}$
 に繋がっている **then**
 36: $V := V - \{v_2\}$
 37: **if** $v_1 < v_2$ **then**
 38: $E' := E' - \{(v_1, v_2)\}$
 39: $A[v_1] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2) \mid (r, \mathcal{X}_1) \in A[v_1],$
 $(r, \mathcal{X}_2, r', \mathcal{X}_3) \in A[(v_1, v_2)], (r', \mathcal{X}_4) \in$
 $A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 = \text{var}(r') \text{ かつ } \mathcal{X}_3 \cap \mathcal{X}_4 =$
 $\emptyset\}$
 40: $B[v_1] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2) \mid (r, \mathcal{X}_1) \in B[v_1],$
 $(r, \mathcal{X}_2, r', \mathcal{X}_3) \in A[(v_1, v_2)], (r', \mathcal{X}_4) \in$
 $A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 = \text{var}(r') \text{ かつ } \mathcal{X}_3 \cap \mathcal{X}_4 =$
 $\emptyset\} \cup \{(r, \mathcal{X}_1 \cup \mathcal{X}_2) \mid (r, \mathcal{X}_1) \in A[v_1],$
 $(r, \mathcal{X}_2, r', \mathcal{X}_3) \in B[(v_1, v_2)], (r', \mathcal{X}_4) \in$
 $B[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 = \text{var}(r') \text{ かつ } \mathcal{X}_3 \cap \mathcal{X}_4 =$
 $\emptyset\} \cup \{(r, \mathcal{X}_1 \cup \mathcal{X}_2) \mid (r, \mathcal{X}_1) \in A[v_1],$
 $(r, \mathcal{X}_2, r', \mathcal{X}_3) \in D[(v_1, v_2)], (r', \mathcal{X}_4) \in$
 $A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 = \text{var}(r') \text{ かつ } \mathcal{X}_3 \cap \mathcal{X}_4 =$
 $\emptyset\}$
 41: **else**
 42: $E' := E' - \{(v_2, v_1)\}$
 43: $A[v_1] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2) \mid (r, \mathcal{X}_1) \in A[v_1],$
 $(r', \mathcal{X}_3, r, \mathcal{X}_2) \in B[(v_2, v_1)], (r', \mathcal{X}_4) \in$
 $A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 = \text{var}(r') \text{ かつ } \mathcal{X}_3 \cap \mathcal{X}_4 =$
 $\emptyset\}$
 44: $B[v_1] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2) \mid (r, \mathcal{X}_1) \in B[v_1],$
 $(r', \mathcal{X}_3, r, \mathcal{X}_2) \in B[(v_2, v_1)], (r', \mathcal{X}_4) \in$
 $A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 = \text{var}(r') \text{ かつ } \mathcal{X}_3 \cap \mathcal{X}_4 =$
 $\emptyset\}$
 else if $v_2 \in V$ が存在し、正確に2つの辺
 $\{v_1, v_2\}, \{v_2, v_3\}$ に繋がっている **then**
 45: $V := V - \{v_2\}$
 46: **if** $v_1 < v_2$ **かつ** $v_2 < v_3$ **then**
 47: $E := E' - \{(v_1, v_2), (v_2, v_3)\}$
 48: $A := \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $A[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in A[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 = \text{var}(r'')$
 $\text{かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 49: $B := \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $B[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in B[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 = \text{var}(r'')$
 $\text{かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 50: $C := \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $C[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in B[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 =$
 $\text{var}(r'') \text{ かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 $\cup \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $A[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in C[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 = \text{var}(r'')$
 $\text{かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 51: $D := \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $B[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in A[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in B[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 =$
 $\text{var}(r'') \text{ かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 $\cup \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $D[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in A[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 =$
 $\text{var}(r'') \text{ かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 52: $E := \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $E[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in B[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 =$
 $\text{var}(r'') \text{ かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 $\cup \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $E[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in E[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 =$
 $\text{var}(r'') \text{ かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 53: $F := \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $F[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in A[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in B[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 =$
 $\text{var}(r'') \text{ かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$
 $\cup \{(r, \mathcal{X}_1, r', \mathcal{X}_2) \mid (r, \mathcal{X}_1, r'', \mathcal{X}_3) \in$
 $F[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in F[(v_2, v_3)],$
 $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 =$
 $\text{var}(r'') \text{ かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}\}$

```

64:  $B[(v_1, v_2)], (r'', \mathcal{X}_4, r', \mathcal{X}_2) \in D[(v_2, v_3)],$ 
65:  $(r'', \mathcal{X}_5) \in A[v_2], \mathcal{X}_3 \cup \mathcal{X}_4 \cup \mathcal{X}_5 = \text{var}(r'')$ 
66:  $\text{かつ } \mathcal{X}_3, \mathcal{X}_4, \mathcal{X}_5 \text{ は互いに素}$ 
67: if  $(v_1, v_3) \notin E'$  then
68:    $E' := E' \cup \{(v_1, v_3)\}$ 
69:    $A[(v_1, v_3)] := A$ 
70:    $B[(v_1, v_3)] := B$ 
71:    $C[(v_1, v_3)] := C$ 
72:    $D[(v_1, v_3)] := D$ 
73: else
74:    $A[(v_1, v_3)] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2, r', \mathcal{X}_3 \cup$ 
75:    $\mathcal{X}_4) \mid (r, \mathcal{X}_1, r', \mathcal{X}_3) \in A[(v_1, v_3)],$ 
76:    $(r, \mathcal{X}_2, r', \mathcal{X}_4) \in C, \mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset \text{ かつ}$ 
77:    $\mathcal{X}_3 \cap \mathcal{X}_4 = \emptyset\} \cup \{(r, \mathcal{X}_1 \cup \mathcal{X}_2, r', \mathcal{X}_3 \cup$ 
78:    $\mathcal{X}_4) \mid (r, \mathcal{X}_1, r', \mathcal{X}_3) \in C[(v_1, v_3)],$ 
79:    $(r, \mathcal{X}_2, r', \mathcal{X}_4) \in A, \mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset \text{ かつ}$ 
80:    $\mathcal{X}_3 \cap \mathcal{X}_4 = \emptyset\}$ 
81:    $B[(v_1, v_3)] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2, r', \mathcal{X}_3 \cup$ 
82:    $\mathcal{X}_4) \mid (r, \mathcal{X}_1, r', \mathcal{X}_3) \in B[(v_1, v_3)],$ 
83:    $(r, \mathcal{X}_2, r', \mathcal{X}_4) \in C, \mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset \text{ かつ}$ 
84:    $\mathcal{X}_3 \cap \mathcal{X}_4 = \emptyset\} \cup \{(r, \mathcal{X}_1 \cup \mathcal{X}_2, r', \mathcal{X}_3 \cup$ 
85:    $\mathcal{X}_4) \mid (r, \mathcal{X}_1, r', \mathcal{X}_3) \in C[(v_1, v_3)],$ 
86:    $(r, \mathcal{X}_2, r', \mathcal{X}_4) \in B, \mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset \text{ かつ}$ 
87:    $\mathcal{X}_3 \cap \mathcal{X}_4 = \emptyset\}$ 
88:    $C[(v_1, v_3)] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2, r', \mathcal{X}_3 \cup$ 
89:    $\mathcal{X}_4) \mid (r, \mathcal{X}_1, r', \mathcal{X}_3) \in C[(v_1, v_3)],$ 
90:    $(r, \mathcal{X}_2, r', \mathcal{X}_4) \in C, \mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset \text{ かつ}$ 
91:    $\mathcal{X}_3 \cap \mathcal{X}_4 = \emptyset\}$ 
92:    $D[(v_1, v_3)] := \{(r, \mathcal{X}_1 \cup \mathcal{X}_2, r', \mathcal{X}_3 \cup$ 
93:    $\mathcal{X}_4) \mid (r, \mathcal{X}_1, r', \mathcal{X}_3) \in D[(v_1, v_3)],$ 
94:    $(r, \mathcal{X}_2, r', \mathcal{X}_4) \in C, \mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset \text{ かつ}$ 
95:    $\mathcal{X}_3 \cap \mathcal{X}_4 = \emptyset\} \cup \{(r, \mathcal{X}_1 \cup \mathcal{X}_2, r', \mathcal{X}_3 \cup$ 
96:    $\mathcal{X}_4) \mid (r, \mathcal{X}_1, r', \mathcal{X}_3) \in C[(v_1, v_3)],$ 
97:    $(r, \mathcal{X}_2, r', \mathcal{X}_4) \in D, \mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset \text{ かつ}$ 
98:    $\mathcal{X}_3 \cap \mathcal{X}_4 = \emptyset\}$ 
99: end if
100: else if  $v_1 < v_2$  かつ  $v_3 < v_2$  then
101:    $E := E' - \{(v_1, v_2), (v_3, v_2)\}$ 
102:   (同様なので省略)
103: else if  $v_2 < v_1$  かつ  $v_3 < v_2$  then
104:    $E := E' - \{(v_2, v_1), (v_3, v_2)\}$ 
105:   (同様なので省略)
106: end if
107: end if
108: end while
109:  $V$  の唯一の要素を  $v$  とする
110: if  $(r, \mathcal{X}) \in B[v]$  が存在して,  $\text{var}(r) = \mathcal{X}$  then
111:   return accept
112: else
113:   return reject
114: end if

```

6 まとめと今後の課題

かつて、創薬の研究は、製薬会社の研究員が製薬会社につながりのある研究者に限られていた。もしくは、膨大な手間と費用とをかけて自前の医薬品データベースを作成しなければならなかった。しかし、最近、自由に利用できるデータベース [5, 11, 12] があいついで公開され、創薬の研究への道が開けている。

本稿では、化学グラフを頂点ラベル付グラフと見なし、化学グラフの木オートマトンによる所属問題を考え、この所属問題が NP 完全であることと、Tree-Width が高々 2 であるような化学グラフに対する線形時間の所属性判定アルゴリズムを提案した。

今後の課題は、木文法を拡張し、化学グラフを生成する文法を提案することである。

参考文献

- [1] Jean-Loup Faulon. Isomorphism, automorphism partitioning, and canonical labeling can be solved in polynomial-time for molecular graphs. *Journal of Chemical Information and*

- Computer Sciences*, Vol. 38, No. 3, pp. 432–444, 1998.
- [2] Eugene M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *J. Comput. Syst. Sci.*, Vol. 25, No. 1, pp. 42–65, 1982.
- [3] László Babai and Eugene M. Luks. Canonical labeling of graphs. In *STOC*, pp. 171–183, 1983.
- [4] John E. Hopcroft and J. K. Wong. Linear time algorithm for isomorphism of planar graphs (preliminary report). In *STOC*, pp. 172–184, 1974.
- [5] ChEMBL. <https://www.ebi.ac.uk/chembl/>.
- [6] Hans L. Bodlaender. Polynomial algorithms for graph isomorphism and chromatic index on partial k-trees. *J. Algorithms*, Vol. 11, No. 4, pp. 631–643, 1990.
- [7] Jirí Matousek and Robin Thomas. On the complexity of finding iso- and other morphisms for partial k-trees. *Discrete Mathematics*, Vol. 108, No. 1-3, pp. 343–364, 1992.
- [8] H. Comon, M. Dauchet, R. Gilleron, F. Jacquemard, D. Lugiez, C. Löding, S. Tison, and M. Tommasi. Tree automata techniques and applications. Available on: <http://www.grappa.univ-lille3.fr/tata>, 2007. release October, 12th 2007.
- [9] Akio Fujiyoshi. Recognition of directed acyclic graphs by spanning tree automata. *Theor. Comput. Sci.*, Vol. 411, No. 38-39, pp. 3493–3506, 2010.
- [10] Stefan Arnborg and Andrzej Proskurowski. Characterization and recognition of partial 3-trees. *SIAM J. Algebraic Discrete Methods*, Vol. 7, No. 2, pp. 305–314, 1986.
- [11] PubChem. <http://pubchem.ncbi.nlm.nih.gov/>.
- [12] Worldwide Protein Data Bank. <http://www.wwpdb.org/>.