

SAT ソルバーを用いた部分グラフ探索のための制約モデル

Constraint Models for SAT-based Subgraph Search

川原 征大*¹ 宋 剛秀*² 番原 睦則*² 田村 直之*²
 Masahiro Kawahara Takehide Soh Mutsumori Banbara Naoyuki Tamura

*¹神戸大学 大学院システム情報学研究科
 Graduate School of System Informatics, Kobe University

*²神戸大学 情報基盤センター
 Information Science and Technology Center, Kobe University

Subgraph search is to find subgraphs satisfying given constraints, and it has various applications such as circuit wiring. In this paper, we discuss constraint models of subgraph search and their SAT encodings to develop a rapid subgraph tool using SAT solvers. In particular, for connectivity constraints, we study pros and cons of two constraint models and propose a new hybrid model of them. Experimental result shows that the hybrid model inherits the advantages of both two models and solves the most number of benchmark instances.

1. はじめに

近年、大規模な命題論理の充足可能性判定 (SAT) 問題を高速に解くことが可能な SAT ソルバーが実現され、プランニング、スケジューリング、制約充足問題など、様々な分野への応用が進んでいる [井上 10] .

部分グラフ探索とは、与えられたグラフから条件を満たす部分グラフを見つける探索のことである。部分グラフ探索は経路探索や回路配線など、実世界に様々な応用があり、重要である。特に、様々なグラフ上の条件を組合せて汎用的に部分グラフ探索を行う方法が研究されている [Inoue 14, Dooms 05] .

本稿では高速な部分グラフ探索を実現するために、SAT ソルバーを用いた部分グラフ探索のための制約モデルとその SAT 符号化について考える。著者らの以前の研究 [川原 15] では、頂点の連結条件を 2 つの制約モデルで表した。推移モデルは頂点の推移関係 [Prestwich 03, Gebser 14] を用いて連結性を定めた制約モデルであり、次数モデルは頂点の次数条件 [Dooms 05] を用いて連結性を定めた制約モデルである。本稿では、まずこれらの制約モデルの SAT 符号化節数について考察を深めその長所と短所を明らかにする。さらに考察結果から 2 つの制約モデルを組み合わせた新しいハイブリッドモデルの提案を行う。ハイブリッドモデルは SAT 符号化した際に、符号化後の節数を少なく抑えることができる点が特長である。

提案方法の評価のために、それぞれ連結条件の性質が異なるマルチパス問題、ハミルトン閉路問題、部分閉路問題の計算機実験を行った。実験の結果、3 つの制約モデルの中でハイブリッドモデルが最も多くの問題を解けることを確認した。

2. 部分グラフ探索

与えられたグラフ $G(\mathcal{V}, \mathcal{E})$ から、条件を満たす部分グラフ $G(\mathcal{V}, E)$ ($E \subseteq \mathcal{E}$) を見つけ出すことを部分グラフ探索という。探索を行う際のグラフ上の条件として、様々なものが考えられる。既存研究の 1 つである Graphillion [Inoue 14] では、以下の条件を指定して、部分グラフ探索を行うことができる。

- ・連結条件 部分グラフの頂点の連結性を指定する。
- ・次数条件 部分グラフの各頂点の次数を指定する。

- ・辺数条件 部分グラフの辺の総数を指定する。
 - ・非閉路条件 部分グラフの閉路の有無を指定する。
- 例えば頂点 s, t を結ぶパスを求める場合、連結条件と次数条件、非閉路条件を指定することで求めることができる。
- ・連結条件 s, t が連結である。
 - ・次数条件 s, t の次数が 1、それ以外の頂点の次数が 0 または 2。
 - ・非閉路条件 部分グラフに閉路を含まない。
- グラフ上の条件 1 つ 1 つは単純であるが、組合せることで経路探索や回路配線など、様々なグラフ上の問題を汎用的に記述することができる。

3. 連結条件の制約モデル

本節ではまず連結条件の定義を行う。次に連結条件を表す 2 つの制約モデル、推移モデル・次数モデルについて説明する。なお、特に混乱が生じない場合は、0-1 変数を命題変数として扱う。例えば、 $a = 1$ および $a = 0$ を、それぞれ a および $\neg a$ として記述することがある。

3.1 連結条件の定義

入力 頂点集合の集合 $\{V_1, \dots, V_m\}$

条件 各 V_i ($1 \leq i \leq m$) 中の全ての頂点は互いに連結

つまり、頂点集合 V 中の全ての頂点が互いに連結である制約を $connected(V)$ と定めると、連結条件は以下のような $connected(V)$ の連言で表せる。

$$\bigwedge_{1 \leq i \leq m} connected(V_i)$$

3.2 準備 (部分グラフ探索の制約表現)

以降で説明する制約モデルで、共通して使用する変数について記述する。求める部分グラフ $G(\mathcal{V}, E)$ の辺集合 $E \in \mathcal{E}$ の特性関数を表す変数 a_e を定義する。

$$a_e \in \{0, 1\} \quad (\forall e \in \mathcal{E})$$

つまり、 $E = \{e \in \mathcal{E} \mid a_e = 1\}$ となる。例えば、頂点 $u \in \mathcal{V}$ の次数が d (d は 0 以上の整数) である、といった次数条件は

連絡先: 川原 征大, 神戸大学 大学院システム情報学研究科, 兵庫県神戸市灘区六甲台町 1-1 神戸大学情報基盤センター 田村研究室, 078-803-5364, kawahara@stu.kobe-u.ac.jp

u に接続する辺の集合 $\mathcal{E}_u = \{e \in \mathcal{E} \mid u \in e\}$ を用いて以下の
ように表現できる .

$$\sum_{e \in \mathcal{E}_u} a_e = d$$

3.3 推移モデル

推移モデルは [Prestwich 03, Gebser 14] の手法を応用した ,
頂点の推移関係を利用したモデルである .

G 上で $V = \{v_1, \dots, v_n\}$ 中の全頂点が互いに連結となる必
要十分条件は , G の部分グラフ G' が以下の条件を満たすこと
である [Gebser 14] (G' は便宜上有向グラフとして考える) .

1. v_1 から v_j ($2 \leq j \leq n$) に向かうパスが存在する .
2. v_1 から $u \in \mathcal{V} \setminus \{v_1\}$ に向かうパスが存在するとき , かつ
そのときに限り u に隣接するある頂点 v について辺 uv
が存在し , v_1 から v に向かうパスが存在する .
3. G' に有向閉路を含まない .

まず , G' を表すために以下の変数を導入する (ただし , $adj(u)$
は G における u の隣接頂点の集合 , $\mathcal{A} = \{(u, v) \mid u \in \mathcal{V}, v \in$
 $adj(v)\}$) .

$$\begin{aligned} a'_{uv} &\in \{0, 1\} & ((u, v) \in \mathcal{A}) \\ r_{uv} &\in \{0, 1\} & (u, v \in \mathcal{V}, u \neq v) \end{aligned}$$

a'_{uv} は G' の辺集合の特性関数を表す変数であり , 各辺に対し
て双方向で定義される . r_{uv} は G' 上で , u から v に向かうパ
スが存在することを表す変数である . 条件 1 , 2 , 3 は以下の
ように制約として表せる .

$$\begin{aligned} r_{v_1 v_j} &= 1 & (2 \leq j \leq n) \\ r_{v_1 u} &\Rightarrow \bigvee_{v \in adj(u)} a'_{vu} & (u \in \mathcal{V} \setminus \{v_1\}) \\ a'_{uv} &\Rightarrow r_{v_1 u} & ((u, v) \in \mathcal{A}) \\ a'_{uv} &\Rightarrow \neg r_{vu} & ((u, v) \in \mathcal{A}) \end{aligned}$$

そして , 頂点間の推移関係を表す制約 [Prestwich 03] を記述
し , a'_{uv} と r_{xy} を正しく対応付ける .

$$\begin{aligned} a'_{uv} &\Rightarrow r_{uv} & ((u, v) \in \mathcal{A}) \\ (a'_{uv} \wedge r_{uv}) &\Rightarrow r_{uv} & ((u, v) \in \mathcal{A}) \end{aligned}$$

最後に , G' が G の部分グラフである条件を記述する .

$$a'_{uv} + a'_{vu} \leq a_e \quad (e \in \mathcal{E}, \{u, v\} = e)$$

以上が推移モデルにおける $connected(V)$ である . これを
 $connected_T(V)$ とする (T は Transition の意) .

3.4 次数モデル

次数モデルは [Dooms 05] の理論を制約モデルとして適用さ
せたモデルである .

まず 2 頂点 s, t の連結性について考える . G 上で s, t が連
結であることの必要十分条件は , G の部分グラフ G' が以下の
条件を満たすことである (G' は有向グラフとして考える) .

1. G' における s の出次数が 1 , 入次数が 0 である .

2. G' における t の出次数が 0 , 入次数が 1 である .

3. G' における $u \in \mathcal{V} \setminus \{s, t\}$ の出次数と入次数が等しく , 0
もしくは 1 である .

まず G' を表すために変数 a'_{uv} を定義する .

$$a'_{uv} \in \{0, 1\} \quad ((u, v) \in \mathcal{A})$$

推移モデルと同様に , a'_{uv} は G' の辺集合の特性関数を表す変
数であり , 各辺に対して双方向で定義される . 条件 1 , 2 , 3 は
以下のように制約として表せる .

$$\begin{aligned} deg_{G'}^+(s) &= 1 \wedge deg_{G'}^-(s) = 0 \\ deg_{G'}^+(t) &= 0 \wedge deg_{G'}^-(t) = 1 \\ deg_{G'}^+(u) &= deg_{G'}^-(u) = 0 \text{ or } 1 \quad (u \in \mathcal{V} \setminus \{s, t\}) \end{aligned}$$

なお , $deg_{G'}^+(u), deg_{G'}^-(u)$ は次のように定義される .

$$\begin{aligned} deg_{G'}^+(u) &= \sum_{v \in adj(u)} a'_{uv} \\ deg_{G'}^-(u) &= \sum_{v \in adj(u)} a'_{vu} \end{aligned}$$

最後に , G' が求める G の部分グラフである条件を記述する .

$$a'_{uv} + a'_{vu} \leq a_e \quad (e \in \mathcal{E}, \{u, v\} = e)$$

以上が G 上で s, t が連結するための制約である . この制約を
 $\delta(s, t)$ とすると , $connected(V)$ ($V = \{v_1, \dots, v_n\}$) は以下の
ような $\delta(s, t)$ の連言で表せる .

$$\delta(v_1, v_2) \wedge \delta(v_2, v_3) \wedge \dots \wedge \delta(v_{n-1}, v_n)$$

これを $connected_D(V)$ とする (D は Degree の意) .

4. SAT 符号化

前節で , 推移モデルと次数モデルの制約モデルを示した . 本
節では , 両モデルの SAT 符号化方法について説明する . さらに
両モデルの SAT 符号化後の節数について考察し , 両モデル
の利点を活かしたハイブリッドモデルの提案を行う .

4.1 SAT 符号化の方針

次数モデルで用いられている次数条件以外の制約について
は , 0-1 変数を命題変数として扱うことで , 容易に節形式へと
変換できる .

次数モデルで用いられている次数条件は基数制約である . 基
数制約は [Sinz 05] 等の方法により SAT 符号化可能である . ま
た , 基数制約を組込み制約として実装している SAT ソルバー
も存在する [Le Berre 10] . 今回用いる基数制約 $\sum_{i=1}^n x_i =$
 k ($x_i \in \{0, 1\}$) の k が取りうる値は最大で 1 と小さい . 今回
は k の値が小さい場合に経験的に性能が良いと知られている
SequentialCounter [Sinz 05] を用いて基数制約の SAT 符号化
を行う .

4.2 節数の比較

$connected_T(V), connected_D(V)$ ($V = \{v_1, \dots, v_n\}$) の
SAT 符号化後の節数をそれぞれ $c_T(V), c_D(V)$ とする .
 $c_T(V), c_D(V)$ は以下のような式で表せる (ただし $P_i =$
 $\{v_i, v_{i+1}\}$, $f(u) = 5|adj(u)| - 3$, $g(u) = 10|adj(u)| - 4$) .

$$c_T(V) = 2|\mathcal{E}||\mathcal{V}| + 2|\mathcal{E}| + |\mathcal{V}| + n - 2$$

$$c_D(V) = \sum_{1 \leq i \leq n-1} \left[3|\mathcal{E}| + \sum_{u \in P_i} f(u) + \sum_{u \in \mathcal{V} \setminus P_i} g(u) \right]$$

特に, $c_D(V)$ の計算式は煩雑である. そこで $c_D(V)$ の値を概算するために, 平均次数 $\mu = 2|\mathcal{E}|/|\mathcal{V}|$ を用いる. つまり, 全ての頂点 $u \in \mathcal{V}$ について, $\text{adj}(u) \simeq \mu$ の近似計算を行う. この $c_D(V)$ の近似計算後の値を $c_D^\mu(V)$ とする. $c_T(V)$ を μ を用いて表した式 $c_T^\mu(V)$, および $c_D^\mu(V)$ は以下のような式となる.

$$c_T^\mu(V) = \mu|\mathcal{V}|^2 + \mu|\mathcal{V}| + |\mathcal{V}| + n - 2$$

$$c_D^\mu(V) = (n-1)\left(\frac{23}{2}\mu|\mathcal{V}| - 10\mu - 4|\mathcal{V}| + 2\right)$$

さらに $c_T^\mu(V)$, $c_D^\mu(V)$ の最高次の項でそれぞれ近似すると, $c_T^\mu(V) \simeq 12n\mu|\mathcal{V}|$, $c_D^\mu(V) \simeq \mu|\mathcal{V}|^2$ となる.

以上より, 連結条件の入力 $\{V\}$ の頂点数 $|V|$ と与えられたグラフ $\mathcal{G}(\mathcal{V}, \mathcal{E})$ の頂点数 $|\mathcal{V}|$ の関係から 2 つの制約モデルの SAT 符号化節の多い少ないが変化することが分かる.

- 概ね $|V| < (1/12)|\mathcal{V}|$ の時に $c_T(V) > c_D(V)$.
- 概ね $|V| > (1/12)|\mathcal{V}|$ の時に $c_T(V) < c_D(V)$.

4.3 ハイブリッドモデルの提案

一般に SAT 問題の難易度の指標として, SAT 符号化後の節数が用いられることが多い. そこで頂点集合制約の SAT 符号化後の節数を削減するために, 推移モデルと次数モデルをハイブリッドしたモデルを提案する. ハイブリッドモデルにおける $\text{connected}(V)$ を $\text{connected}_H(V)$ とする. $\text{connected}_H(V)$ は以下のように振る舞う.

$$\text{connected}_H(V) = \begin{cases} \text{connected}_T(V) & (c_T(V) \leq c_D(V)) \\ \text{connected}_D(V) & (c_T(V) > c_D(V)) \end{cases}$$

ハイブリッドモデルにおける連結条件は以下のように表せる.

$$\bigwedge_{1 \leq i \leq m} \text{connected}_H(V_i)$$

5. 計算機実験

ここまで説明した推移モデル, 次数モデル, ハイブリッドモデルを SAT ソルバーを用いた部分グラフ探索ツール Scarabellion [川原 15] 上に実装して, 求解性能の評価を行った. SAT ソルバーは Sat4j [Le Berre 10] を使用した. 全ての実験は Intel Xeon 3.50GHz を搭載した計算機を用いて行った.

5.1 使用したベンチマーク

各制約モデルの性能を評価するために, 以下の 3 種類のベンチマークを使用した. それぞれ, 与えられたグラフ $\mathcal{G}(\mathcal{V}, \mathcal{E})$ と連結条件の入力 $\{V_1, \dots, V_m\}$ の関係を示す.

ベンチマーク	$\mathcal{G}(\mathcal{V}, \mathcal{E})$ と $\{V_1, \dots, V_m\}$ の関係
マルチパス問題	$ V_i = 2 \ (1 \leq i \leq m)$
ハミルトン閉路問題	$ V_i = \mathcal{V} \ (m = 1)$
部分閉路問題	$(1/64) \mathcal{V} \leq V_i \leq \mathcal{V} \ (m = 1)$

節 4.2 で説明した式と $\mathcal{G}(\mathcal{V}, \mathcal{E})$ と $\{V_1, \dots, V_m\}$ の関係から, マルチパス問題では次数モデルの SAT 符号化節数が推移モデルより少なくなり, ハミルトン閉路問題では推移モデルの SAT 符号化節数が次数モデルより少なくなる. 部分閉路問題は次数モデルと推移モデルによる SAT 符号化節数の多い少ないが切り替わる問題になる. 提案するハイブリッドモデルは常に SAT 符号化節数が少ない制約モデルを選択する.

表 1: マルチパス問題: 解けた問題数の比較

	推移モデル	次数モデル	提案方法
解けた問題数	4	9	9

表 2: ハミルトン閉路問題: 解けた問題数の比較

頂点数	問題数	推移モデル	次数モデル	提案方法
1 - 100	23	23	16	23
101 - 200	21	10	0	10
201 - 500	34	7	0	7
≥ 501	29	0	0	0
合計	107	40	16	40

マルチパス問題とハミルトン閉路問題は論文 [川原 15] と同じものを用いており, それぞれ 11 問と 107 問のベンチマークで構成される.

部分閉路問題は今回の実験のために作成したハミルトン閉路問題の亜種であり, 与えられた頂点集合を含む閉路を求める問題である. 部分閉路問題の入力は部分集合 $V \subseteq \mathcal{V}$ (\mathcal{V} は与えられたグラフの頂点集合) であり, 以下 2 つの条件を満たす部分グラフを求める問題となる.

1. 連結条件として V 中の全ての頂点は互いに連結
2. 次数条件として頂点 $u \in V$ の次数が 2, それ以外の頂点の次数が 0 もしくは 2

連結条件は V を用いて $\text{connected}(V)$ になる. $V = \mathcal{V}$ の場合はハミルトン閉路問題となる. ハミルトン閉路問題の実験で使用した 107 個のグラフの中で, 頂点数が 128 以上のもの 77 個をベースに問題を生成した. 具体的には, 各 77 個のグラフに対して部分頂点集合の要素数が $|V| = (1/64)|\mathcal{V}|, \dots, (7/64)|\mathcal{V}|, (1/8)|\mathcal{V}|, \dots, (7/8)|\mathcal{V}|, |\mathcal{V}|$ を満たすように計 15 セットのベンチマークを生成して実験を行った. なお V の要素となる頂点は与えられたグラフからランダムに選択した.

5.2 実験結果

まず上述のベンチマークに対して, 推移モデルと次数モデルによって実際に生成される SAT 符号化節数を調べたところ節 4.2 で説明した $c_T(V)$ 及び $c_D(V)$ の計算式と一致しており, 2 つの制約モデルを正しく実装できていることを確認した.

次に求解性能の比較を行った. マルチパス, ハミルトン閉路, 部分閉路問題の実験結果を表 1, 2, 3 に示す.

マルチパス問題 (表 1) では, 次数モデルが 11 問中 9 問と, 推移モデルに比べて多くの問題を解くことができ, 推移モデルで解けた全て問題は次数モデルより高速に求解できていた. また全ての問題において, 実際に $c_D(V) < c_T(V)$ であり, ハイブリッドモデルは次数モデルを選択している.

ハミルトン閉路問題 (表 2) では, 推移モデルが 107 問中 40 問と, 次数モデルに比べて多くの問題を解くことができ, 次数モデルで解けた全て問題は推移モデルより高速に求解できていた. また全ての問題で, 実際に $c_D(V) > c_T(V)$ であり, ハイブリッドモデルは推移モデルを選択している.

部分閉路問題 (表 3) では, 入力となる部分頂点集合が $|V| \leq (4/64)|\mathcal{V}|$ の時に次数モデルが推移モデルより良い性能を示した. 推移モデルは $(4/64)|\mathcal{V}|$ を境に徐々に次数モデルより求解数が増える結果となった. ハイブリッドモデルは $(6/64)|\mathcal{V}|$ の

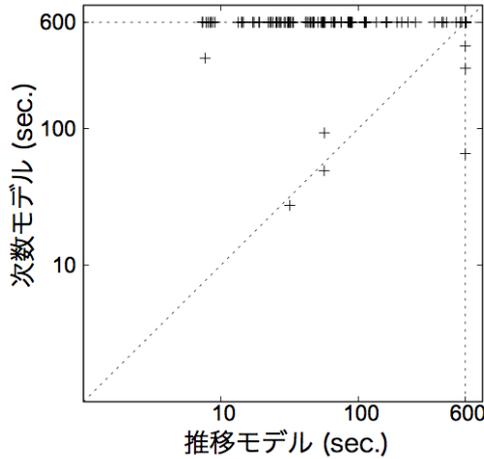


図 1: 推移モデルの節数が次数モデルより少なくなる問題における求解時間の比較

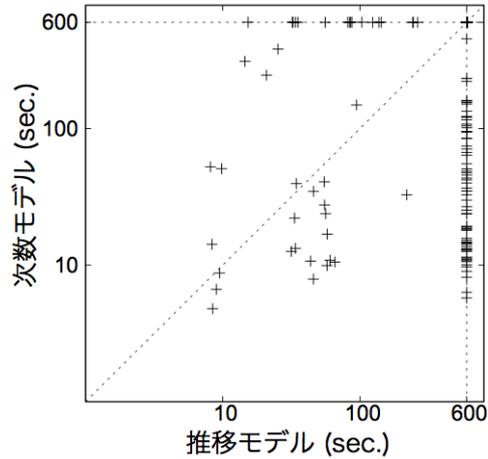


図 2: 次数モデルの節数が推移モデルより少なくなる問題における求解時間の比較

表 3: 部分閉路問題の求解数

入力 V の要素数	問題数	推移モデル	次数モデル	提案方法
$(1/64) V $	77	5	33	33
$(2/64) V $	77	7	28	28
$(3/64) V $	77	7	20	20
$(4/64) V $	77	7	11	11
$(5/64) V $	77	9	7	7
$(6/64) V $	77	8	5	4
$(7/64) V $	77	7	3	7
$\geq (1/8) V $	616	98	3	98
合計	1155	148	110	208

前後で次数モデル・推移モデルの選択が切り替わり、最も多くの問題を解くことに成功した。

今回は提案するハイブリッドモデルの制約モデル選択の指標に節数を用いたが、これが妥当かどうか以下で考察する。図 1, 2 は部分閉路問題における次数モデル, 推移モデルの求解時間を比較した散布図である。横軸は推移モデルの求解時間, 縦軸は次数モデルの求解時間であり, 少なくとも片方のモデルが解けた問題に対してプロットを行った。図 1 の推移モデルの節数が次数モデルより少なくなる問題では, 108 問中 103 問 (95%) において推移モデルがより良い性能を示した。また図 2 の次数モデルの節数が推移モデルより少なくなる問題では 120 問中 95 問 (79%) において次数モデルがより良い性能を示した。これらの結果から, 節数を指標にして切り替える方法は, 少なくとも部分閉路問題に対しては, 概ね良い選択を行うことが分かる。

6. まとめ

本稿では SAT ソルバーを用いた部分グラフ探索方法のための制約モデルについて考察した。具体的には, 頂点の連結条件について, グラフ $G(V, E)$ と頂点集合 V が与えられた時に推移モデルと次数モデルの SAT 符号化節数を計算する式を明らかにした。この式より概ね $|V| < (1/12)|V|$ のときに次数モデルの節数が推移モデルより少なくなり, 概ね $|V| > (1/12)|V|$ のときに推移モデルの節数が次数モデルより少なくなることを明らかにした。さらに, 常に SAT 符号化節数が少なくなるよ

うに制約モデルを選択するハイブリッド符号化の提案を行った。マルチパス, ハミルトン閉路, 部分閉路問題を用いた計算機実験の結果, 節数の多い少ないと連動して推移・次数モデルの求解性能に違いが出ることを, 提案したハイブリッドモデルが適切に推移・次数モデルを切り替え, 最も多くの問題を求解できることを確認した。

参考文献

- [Dooms 05] Dooms, G., Deville, Y., and Dupont, P.: CP (graph): Introducing a graph computation domain in constraint programming, in *Principles and Practice of CP 2005*, pp. 211–225, Springer (2005)
- [Gebser 14] Gebser, M., Janhunen, T., and Rintanen, J.: SAT modulo graphs: Acyclicity, in *Logics in Artificial Intelligence*, pp. 137–151, Springer (2014)
- [井上 10] 井上 克巳, 田村 直之: SAT ソルバーの基礎, 人工知能学会誌, Vol. 25, No. 1, pp. 57–67 (2010)
- [Inoue 14] Inoue, T., Iwashita, H., Kawahara, J., and Minato, S.: Graphillion: software library for very large sets of labeled graphs, *International Journal on Software Tools for Technology Transfer*, pp. 1–10 (2014)
- [Le Berre 10] Le Berre, D. and Parrain, A.: The Sat4j Library, release 2.2, *Journal on Satisfiability, Boolean Modeling and Computation*, Vol. 7, No. 2-3, pp. 59–64 (2010)
- [Prestwich 03] Prestwich, S.: SAT problems with chains of dependent variables, *Discrete Applied Mathematics*, Vol. 130, No. 2, pp. 329–350 (2003)
- [Sinz 05] Sinz, C.: Towards an optimal CNF encoding of boolean cardinality constraints, in *Principles and Practice of CP 2005*, pp. 827–831, Springer (2005)
- [川原 15] 川原 征大, 宋 剛秀, 番原 睦則: SAT ソルバーを用いた高速な部分グラフ探索ツールの実装と評価, 日本ソフトウェア学会大会論文集, Vol. 32, p. 9p (2015)