

フーリエ変換を用いた命題論理式の充足可能性に関する一考察

Analyzing Satisfiability of Propositional Formulas with the Fourier Transform

宮城 智輝*¹ 山本 泰生*² 岩沼 宏治*²
 Tomoki Miyagi Yoshitaka Yamamoto Koji Iwanuma

*¹山梨大学大学院医学工学総合教育部コンピュータ・メディア工学専攻

Department of Computer Science and Media Engineering,
 Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi

*²山梨大学大学院医学工学総合研究部

Department of Research Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi

This paper empirically analyzes satisfiability (SAT) of propositional formulas (PF) in the context of the Fourier transform. The spectral representation of PFs is given as the linear combination of each base function and its Fourier coefficient. In this paper, we first show the equivalence between SAT and existence of non-zero coefficients. In terms of SAT, it is well-known that there is a phase transition as the ratio of the number of clauses to variables in a formula. We next show an empirical result that there is no such a transition in terms of the number of non-zero coefficients, called t -value, in randomly generated problems. This report then investigates the distribution of t -values over the SAT problems as well as the relation with the number of (minimal) models. Finally, we report a feature that there is some biased appearance of t -values, and discuss about it using Hadamard matrixes.

1. はじめに

SAT 問題とは、与えられた命題論理式の充足可能性 (*Satisfiability*) を判定する問題である。最初期に NP 完全性が証明された問題として知られているが、近年、SAT 問題を解く SAT ソルバの高速化が進み、数百万変数から成る実問題でも求解可能であることが報告されている [1]。

これに対して、ランダムに生成された (k -)SAT 問題を用いて、統計的に問題の振る舞いを解析する研究も行われている。ランダム問題の充足可能性は変数の個数に対する節数の比 (制約密度) に依存することが知られている [2]。一般に制約密度の値が高くなるにしたがい、問題が充足可能となる確率は低くなるが、特に制約密度が 4.3 付近でその確率が急激に低下する相転移現象 (*phase transition*) が確認されている。この現象は熱力学などにおいても観測される現象であるが、SAT 問題において相転移現象が生じる理由やメカニズムについてはまだよくわかっていない。

本研究では、このようなランダム問題における充足可能性の特徴について、ブール関数のフーリエ変換をもとに考察を行う。ブール関数のフーリエ表現は、各基底関数とそのフーリエ係数の線形結合で与えられるスペクトル表現をさす。この表現はブール関数の学習 [4] や関連変数の抽出 [5] などの問題に利用されてきている。その一方で、ランダムに作られた SAT 問題の振る舞いをフーリエ表現を用いて調査する研究は著者らの知る限りこれまでなされていない。

そこで本稿では、ブール関数のフーリエ表現と命題論理式の充足可能性に関する関係を明らかにし、ランダム問題を用いてフーリエ表現の見地から SAT 問題の振る舞いを解析する。はじめに、フーリエ表現における非ゼロなフーリエ係数の存在と命題論理式の充足可能性が等価であることを示す。これは充足可能性と (非ゼロ) フーリエ係数との結びつきを意味する。

そこで次に、フーリエ表現のもとで SAT 問題の相転移現象を考察するため、ランダム 3-SAT 問題に対して、非ゼロフーリエ係数の個数 (以降、 t 値と呼ぶ) と制約密度の相関を調査した。その結果、充足可能な問題と不可能な問題を含むすべての問題群では、制約密度の増加に伴い、 t 値の平均が減少する傾向が確認された。その一方で、充足可能な問題群に関しては、制約密度を変化させても t 値の平均が一定値を取る傾向にあった。すなわち、充足可能な問題群に関しては、 t 値は制約密度に影響しなかった。この原因を明らかにするため、次に、 t 値と SAT 問題のモデル数との相関を調査した。モデルは一般に、充足可能となる全変数に対する割り当てと定義されるが、本稿では、いくつかの極小モデルの概念を導入し、それぞれの極小モデルに対して、 t 値との相関を調査した。その結果、 t 値の分布には、大きく 2 つの範囲に偏る傾向があり、 t 値とモデル数の間に中程度の正の相関関係があることがわかった。本稿では最後に、このような t 値の出現傾向が見られる原因について、Hadamard 行列を用いて考察を行う。

論文の構成は次の通りである。2 章で準備として SAT 問題とブール関数のフーリエ変換の各定義を示し、3 章でフーリエ係数と充足可能性の関係を示す。4 章でフーリエ係数と極小モデルの関係を示し、5 章で t 値に関する議論を通して、 t 値の出現に偏りが見られる原因について理論的な考察を試みる。最後に 6 章で本稿をまとめる。

2. 準備

2.1 SAT 問題

命題論理の充足可能性判定 (*propositional satisfiability (testing)*; SAT) は与えられた命題論理式の充足可能性を判定する問題である [7, 8]。SAT の各インスタンスは SAT 問題 (SAT problem instance) と呼ばれる。

ある SAT 問題に出現する命題変数を x_1, x_2, \dots, x_n で表し、命題変数の集合を V とする。このとき、 V の各命題変数は 1 または 0 の値を取り、それぞれ真 (*true*) または偽 (*false*) に対応する。 V に対する (真偽値) 割当ては関数 $A: V \rightarrow \{1, 0\}^n$ と

連絡先: 宮城 智輝, 山梨大学大学院医学工学総合教育部コンピュータ・メディア工学専攻, 住所: 〒400-8511 山梨県甲府市武田 4-3-11 E-mail: g12mk028@yamanashi.ac.jp

定義される。リテラル (*literal*) はある命題変数またはその否定である。節 (*clause*) はリテラルの選言である。節の連言は連言標準形 (*conjunctive normal form; CNF*) と呼ばれる。SAT問題は、通常 CNF 式の論理式で与えられる。また、リテラルの連言を項 (*term*) といい、項 π がブール関数 f の主項であるとは、任意の t の部分項 $t' \subseteq t$ に対して、 $t' \models f$ ならば $t' = t$ が成り立つときまたそのときに限る。

項の選言は選言標準形 (*disjunctive normal form; DNF*) と呼ばれる。命題論理式の DNF 式は、一般にその式のモデルをすべて列挙した表現形式として利用される。本稿では以下の 3 つの DNF 式を用いる。

定義 1 (主項 DNF 式) DNF 式 Ψ の任意の項 π と π 中の任意のリテラル l に対して以下の式が成り立つとき、 Ψ は主項 DNF 式という。

$$\Psi \neq (\Psi - \{\pi\}) \cup \{\pi - \{l\}\}$$

定義 2 (必須項 DNF 式) DNF 式 Ψ の任意の項 π に対して以下の式が成り立つとき、 Ψ は必須項 DNF 式という。

$$\Psi \neq (\Psi - \{\pi\})$$

定義 3 (必須主項 DNF 式) DNF 式 Ψ が主項 DNF かつ必須項 DNF であるとき、 Ψ は必須主項 DNF 式という。

主項 DNF 式、必須項 DNF 式、必須主項 DNF 式のいずれも論理的に等価であり、式を構成する各項は、一つのモデルとみなせる。ただし、一般のモデルは全変数に対する割当てであるのに対して、上記の項はある部分変数に対する割当てとなる。すなわち、上記 3 つの項はそれぞれ異なる極小モデルと考えることができる。

φ をある CNF 式、 V を φ 中に出現する変数集合とし、 A を V に対する割当てとする。割当て A により φ が真となるとき、 A は φ を充足する (*satisfy*) といい、 A は φ のモデル (*model*) であるという。CNF 式 φ が充足可能 (*satisfiable*) であるとは、 φ のモデルが存在することを意味し、充足不能 (*unsatisfiable*) であるとは、 φ のモデルが存在しないことを意味する。

2.2 ブール関数のフーリエ変換

n 変数のブール関数 f において、 f の期待値を $E[f]$ で表す。すなわち、 $E[f] = \frac{1}{2^n} \sum_{\vec{x} \in \{0,1\}^n} f(\vec{x})$ である。次に、 $\chi_{\vec{z}}(\vec{x}) = (-1)^{\vec{z} \cdot \vec{x}}$ を基底ベクトル $\vec{z} \in \{0,1\}^n$ に対する基底関数と呼ぶ。この関数は $\vec{u}, \vec{v} \in \{0,1\}^n$ で $\vec{u} \neq \vec{v}$ ならば $E[\chi_{\vec{u}} \chi_{\vec{v}}] = 0$ を満たす。またすべての \vec{z} について $E[\chi_{\vec{z}}^2] = 1$ が成り立つ。ある基底ベクトル \vec{z} に対する f のフーリエ係数を、 $\hat{f}(\vec{z}) = E[f \chi_{\vec{z}}]$ と定義する。このとき $f(\vec{x})$ は $f(\vec{x}) = \sum_{\vec{z} \in \{0,1\}^n} \hat{f}(\vec{z}) \chi_{\vec{z}}(\vec{x})$ と一意に表され、このスペクトル表現への変換を f のフーリエ変換と呼ぶ。

2.3 関連変数

これまで、ブール関数におけるフーリエ変換はブール関数の学習などに利用されてきた [4, 6, 3]。ブール関数の学習では、関数の出力値に影響を与える変数に注目することが、学習効率の面から重要である。以下では関数の出力値に影響を与える変数の定義を行う。

ある割当て $A \in \{0,1\}^n$ において、1 箇所だけ A と割当ての異なる A' を A の隣接インスタンスと呼ぶ。すなわち、 $\vec{\alpha} \in \{0,1\}^{k-1}$ 、 $\vec{\beta} \in \{0,1\}^{n-k}$ において $\vec{\alpha} \circ 0 \circ \vec{\beta}$ は $\vec{\alpha} \circ 1 \circ \vec{\beta}$ の隣接インスタンスである。特に $k = i$ ならば、その 2 つを i 番目の変数 x_i における隣接インスタンスと呼ぶ。この隣接インスタンスを用いて関連変数は以下のように定義される。

定義 4 (関連変数) $\{0,1\}^n$ を入力とする関数 f について、 i 番目の変数 x_i における隣接インスタンス A, A' において $f(A) \neq f(A')$ となる A, A' が存在するとき、 x_i を f の関連変数という。

また関連変数とフーリエ係数の間には以下の補題 [5] が成り立つ。

補題 1 $f : \{0,1\}^n \rightarrow \mathbb{R}$ において $x_i (1 \leq i \leq n)$ が関連変数でないとき、かつそのときのみ、すべての $\vec{\alpha} \in \{0,1\}^{i-1}$ 、 $\vec{\beta} \in \{0,1\}^{n-i}$ について $\hat{f}(\vec{\alpha} \circ 1 \circ \vec{\beta}) = 0$ である。

3. フーリエ係数と充足可能性の関係

ブール関数 f のフーリエ係数と関連変数の間には以下の命題が成り立つ。

命題 1 恒真でないブール関数 f が関連変数を持つならば、 f は充足可能である。

これらの補題と命題を用いると以下のフーリエ係数と充足可能性に関する以下の関係が成り立つ。

定理 1 恒真でないブール関数 f において、非ゼロフーリエ係数が存在するときまたそのときに限り、 f は充足可能である。

証明 1 恒真でないブール関数 f において、 $\hat{f}(\vec{z}) \neq 0$ となる基底ベクトル \vec{z} が存在するとき、補題 1 の対偶より、ある x_i が関連変数であるから、命題 1 より、恒真でないブール関数 f は充足可能と判定される。また、恒真でないブール関数 f が充足可能であれば、命題 1 より、ある x_i が関連変数であるから、補題 1 の対偶より、非ゼロなフーリエ係数 $\hat{f}(\vec{\alpha} \circ 1 \circ \vec{\beta}) \neq 0$ が存在する。□

この定理は、充足可能性と非ゼロフーリエ係数との強い結びつきを意味する。そこで次に、フーリエ表現をもとに SAT 問題の相転移現象を解析することを目的として、 t 値と呼ばれる非ゼロなフーリエ係数の個数と制約密度の関係性を検討する。まずはじめに、10 変数のランダム 3-SAT 問題の各制約密度 (1.0 ~ 10.0; 0.5 刻み) に対する充足可能な問題の割合の推移を図 1 に示す。

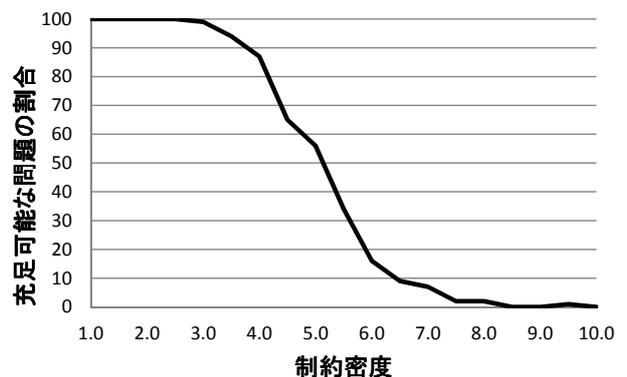


図 1: 各制約密度に対する充足可能な問題の割合の推移

図 1 の縦軸は 100 問のうち充足可能となる問題の割合を示し、横軸は制約密度を示す。図 1 の結果から制約密度の増加とともに、充足可能な問題の割合が減少する相転移現象とみなせる推移が確認された。

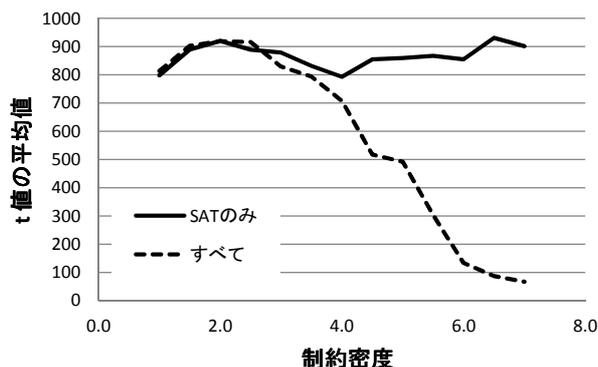


図 2: 各制約密度に対する t 値の平均値の推移

次に、このような推移を示す問題群の各制約密度に対する t 値の平均値の推移を図 2 の破線で示す。

図 2 の縦軸は t 値の平均値を示し、横軸は制約密度を示す。この結果から制約密度の増加にともない、t 値の平均値が減少する傾向が確認された。特に、制約密度 2.0 以降で t 値の平均値は減少し、制約密度が 4.5 で t 値の平均値は t 値の上限の約半分の値になることが確認された。また、制約密度 6.0 以降は t 値の平均値は低い値をとることが確認された。以上の結果は、制約密度の増加にともない、t 値が 0 を示す充足不能と判定される問題の割合が増加したことが原因であると推測される。そこで、次に充足可能と判定される問題のみに対して、t 値の平均値と制約密度の推移を求めた。実験結果を図 2 の実線により示す。結果より、t 値の平均値と制約密度の間に関係性は確認されなかった。

4. フーリエ係数と極小モデルの関係

4.1 極小モデルの定義

そこで本章では、充足可能と判定される問題群に対して、t 値の平均値と制約密度の相関が確認されなかった原因を明らかにすることを目的として、制約密度と相関を持つモデル数に着目し、t 値とモデル数の関係性を調査する。全変数に対する割当てと定義される一般のモデルだけでなく、本研究では、極小モデルとして、2 章で定義した主項 DNF 式、必須項 DNF 式、必須主項 DNF 式の 3 種類を用いる。なおそれぞれの DNF 式は、与えられた CNF 式から、文献 [9] の双対化計算法を用いて導出した。

4.2 実験

本節では項の数とフーリエ係数の関係性を検証する。まず初めに、3 章で用いた充足可能となる問題群に対して t 値とモデルの分布を調べて、以下に定義する相関係数 r を用いて相関を求めた。実験の結果を図 3 に示す。

定義 5 (相関係数) 変数 x, y のデータが n 組あるとき、相関係数 r は以下の式で与えられる。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

図 3 の縦軸は t 値を示し、横軸はモデルを示す。図 3 の結果から、プロットした点は 2 つの範囲に偏る傾向が確認された。1 つは t 値が上限の $t = 1024$ の範囲であり、もう 1 つは t 値が $t = 910$ 以下の範囲である。また、相関係数の値は $r = 0.450$

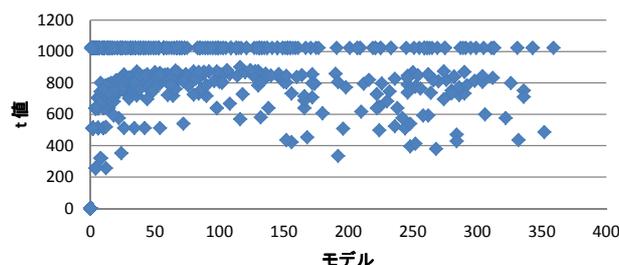


図 3: t 値とモデルの分布

であった。この結果より、t 値とモデルの間には中程度の正の相関関係があり、モデルが多いほど t 値が高くなる傾向が確認された。

次に、主項の数と t 値の分布を調べて相関を示す。実験の結果を図 4 に示す。図 4 の縦軸は t 値を示し、横軸は主項の項数

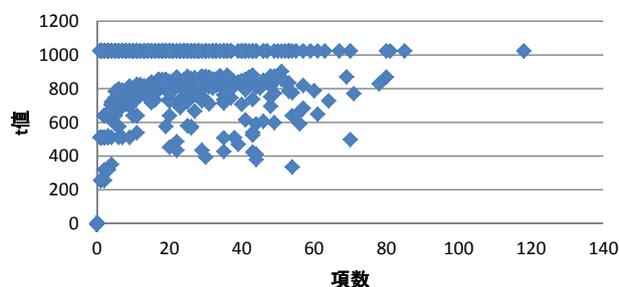


図 4: t 値と主項の項数の分布

を示す。図 4 の結果に関しても、プロットした点は 2 つの範囲に偏る傾向が確認された。また、相関係数の値は $r = 0.563$ であった。この結果により t 値と主項の項数との間にも中程度の正の相関関係があり、項数が多いほど t 値が高くなる傾向が確認された。

次に、必須項の数と t 値の分布を調べて相関を示す。実験結果を図 5 に示す。図 5 の縦軸は t 値を示し、横軸は必須項の項

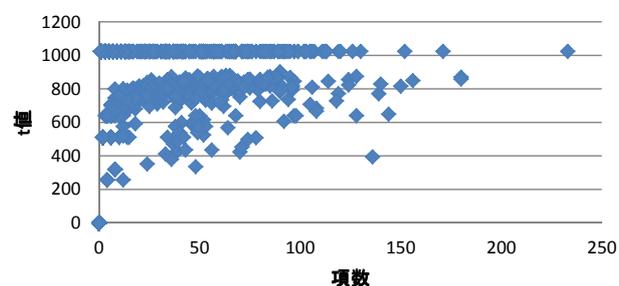


図 5: t 値と必須項の項数の分布

数を示す。図 5 の結果に関しても、プロットした点は 2 つの範囲に偏る傾向が確認された。また、相関係数の値は $r = 0.572$ であった。また、この結果より t 値と必須項の項数との間には中程度の正の相関関係があり、項数が多いほど t 値が高くなる傾向が確認された。

次に、必須主項の数と t 値の分布を調べて相関を示す。実験結果を図 6 に示す。図 6 の縦軸は t 値を示し、横軸は主項の項

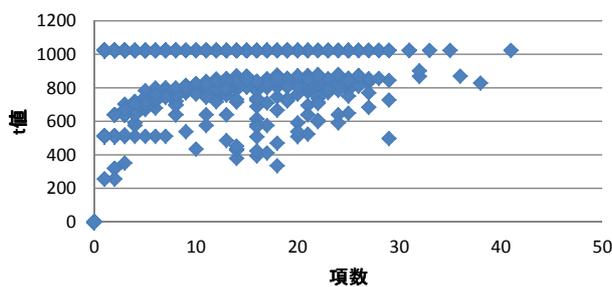


図 6: t 値と必須主項の項数の分布

数を示す。図 6 の結果に関しても、プロットした点は 2 つの範囲に偏る傾向が確認された。また、相関係数の値は $r = 0.658$ であった。また、この結果より t 値と必須主項の項数との間にも中程度の正の相関関係があり、項数が多いほど t 値が高くなる傾向が確認された。

5. t 値に関する議論

これまでの実験を通して、ランダム問題に対して、 t 値の出現分布にはある偏りが見られることがわかった。すなわち t 値は、(10 変数に対して) 取りうる最大値 1024 からせいぜい 200 までの値を層を成して分布しており、比較的低い値を取ることはなかった。

フーリエ変換を用いた関連変数の発見法において、 t 値はその計算量を決定付ける重要なパラメータの一つであることが知られている [5]。その点で、実験結果が示す t 値の特徴が理論的にも成立するか検討することは意義深いものと考えられる。非ゼロフーリエ係数の値については、文献 [4] において、 f の次数との興味深い性質が明らかになっている。すなわち f は、極端に大きな次数と極端に大きなフーリエ係数を同時に取りえないという性質である。これに対して、非ゼロフーリエ係数の個数である t 値は、次数といったブール関数の構成的特性ではなく、フーリエ係数を記述する際に利用される Hadamard 行列 [3] の特性と密接に関わっている。

以下のような、Sylvester 型の Hadamard 行列を考える。

$$\begin{aligned} H_1 &= \begin{bmatrix} 1 \end{bmatrix}. \\ H_2 &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \\ H_{2^{n+1}} &= \begin{bmatrix} H_{2^n} & H_{2^n} \\ H_{2^n} & -H_{2^n} \end{bmatrix}. \end{aligned}$$

各基底ベクトル \vec{x}_i ($1 \leq i \leq 2^n$) のフーリエ係数 $\hat{f}_{\vec{x}_i}$ から成るベクトルを $A(f)$ とし、 \vec{x}_i ($1 \leq i \leq 2^n$) の f の出力値 $f(\vec{x}_i)$ から成るベクトルを \vec{f} とする。このとき、 $A(f) = \frac{1}{2^n} H_{2^n} \vec{f}$ が成り立つことが知られている [3]。

\vec{f} は 0-1 ベクトルであり、 f のモデル数は \vec{f} 中の 1 の要素の個数に一致する。Hadamard 行列の要素はすべて非ゼロ値を取るため、モデル数が 1 である任意の f の t 値は、必ず 2^n となることがわかる。同様にモデル数が 2 である任意の f の t 値は、各 Hadamard 列が満たす直交性の性質から 2^{n-1} となる

ことがわかる。本稿で行った実験結果では、 t 値はモデル数に依存せず、比較的小さな値を取っていたが、Hadamard 行列の観点からもこのような理論的考察を進めることが可能である。

6. まとめ

本論文では、フーリエ表現におけるフーリエ係数とブール関数の充足可能性の関係性について考察を行った。その結果、ランダム 3-SAT 問題に対しては制約密度の増加に伴い、 t 値が減少する傾向が確認されたが、充足可能な 3-SAT 問題に対しては t 値と制約密度の間に相関性は確認されなかった。また、フーリエ係数と極小モデルの相関性を実験により確かめたところ、 t 値とモデル及び項数ともに中程度の正の相関関係があり、モデルや項数が多いほど、 t 値が高くなる傾向が確認された。また、 t 値とモデル、 t 値と項数ともに値の分布が 2 つの層に分かれる性質が見られた。

この t 値の出現傾向に関する理論的な考察は本論文では明らかにできなかったが、アダマール行列の性質と何らかの関係があると予測される。今後は t 値と Hadamard 行列の関係性に着目し、 t 値の理論的な特性をさらに解析する予定である。

謝辞

本研究は一部、文科省科学研究費補助金 (若手 B: No.22700141) および文科省科学研究費補助金 (基盤 C: No.22500127) の援助を受けている。

参考文献

- [1] Cook, S.: The complexity of theorem-proving procedures, *Proc. the 3rd Annual ACM Symp. on Theory of Computing (STOC'71)*, pp. 151-158, ACM (1971)
- [2] Mitchell, D. G., Selman, B.: Hard and easy distributions of SAT problems. *Proc. AAAI-92*, pp. 459-465 (1992).
- [3] Bruck, J.: Chapter 12 in the book "Boolean Models and Methods in Mathematics, Computer Science, and Engineering". Cambridge University Press, pp. 531-553 (2010)
- [4] Mossel, E., O'Donnell, R.: Learning Juntas. *Computer Science Department, Paper1180*, pp. 1-13 (2004)
- [5] 飯田 雅臣: フーリエ変換を用いた関連変数の発見法, 電子情報通信学会 信学技報, Vol. 101, No. 431, pp. 43-50, (2001)
- [6] 天野 一幸, 瀧本 英二: ブール関数のフーリエ変換とその応用, 電子情報通信学会誌, Vol. 82, No. 12, pp. 1270-1272 (1999)
- [7] 鍋島 英知, 宋剛秀: 高速 SAT ソルバーの原理, 人工知能学会誌, Vol. 25, No. 1, pp. 68-76 (2010)
- [8] 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)
- [9] 山本 泰生, 鍋島 英知, 岩沼 宏治: 一般双対化問題における冗長節生成の抑制法とその評価. 第 25 回人工知能全国大会, 3J1-OS7-9 (2011)