3充足可能性判定問題3SATの正例題生成手法の解析

望月 厚[†] 元木光雄[††] 渡辺 治[†††]

NP完全問題の1つである3充足可能性判定問題3SATに対し，その正の例題（ならびに，その充足解）をランダムに生成する手法—例題生成アルゴリズム—を解析する．例題生成アルゴリズムには，「解が1つに定まる例題だけを生成する」という特徴が要求される場合がある．本論文では，ある単純な例題生成アルゴリズムを取り上げ，そのアルゴリズムで変数 n，項数 m の3SATの要素を生成したとき，解が1つしかない例題が生成される条件を調べた．理論的には，$m \geq d_1 n^2$（ただし d_1 はある正定数）のとき，高い確率で解が1つしかない例題が生成されることを示した．一方，実験的には，$m \geq d_2 n \log n$（ただし d_2 はある正定数）のとき，高い確率で解が1つしかない例題が生成されるという結果を得た．

On a Positive Instance Generation for the 3-Satisfiability (3SAT) Problem

Atsushi Mochizuki,[†] Mitsuo Motoki[††] and Osamu Watanabe[†††]

For the 3-Satisfiability Problem (3SAT), we consider a problem of generating its positive instance (and its solution) randomly. For such a problem, it is desirable in some cases if an instance generation algorithm produces only (and all) "unique solution instances", i.e., instances with only one solution. In this paper, we consider one simple instance generation algorithm, and investigate a condition that the algorithm generates unique solution instances when generating 3SAT instances with n variables and m clauses. We first prove that if $m \geq d_1 n^2$ (where d_1 is some constant), then the algorithm generates unique solution instances with high probability. We also report some experimental result showing that if $m \geq d_2 n \log n$ (where d_2 is some constant), then the algorithm generates unique solution instances with high probability.

1. はじめに

本論文では，NP問題の正例題生成問題を考える．具体的には，3充足可能性判定問題3SATの正例題生成手法を1つ取り上げ，その解析を行う．

NP問題の正例題生成問題とは，NP集合 X に対し，X の元 x と $x \in X$ の証拠（ふつう「x の解」と呼ばれている）を，ランダムに生成する問題である．本論文で取り上げる3SATの場合，集合3SAT（充足可能な3CNF論理式の集合）に入る論理式 f と，その f に対する充足割当 t をランダムに生成する問題が，3SATの正例題生成問題であり，そのためのアルゴリズムが，3SATの正例題生成手法である．

一般に，例題生成手法は，アルゴリズムのテストや性能評価に用いられる．NP問題の正例題生成問題も，NP問題に対する様々な発見的手法の有効性などを確かめるために使うことができる．また，平均的に難しいと予想されているNP問題の正例題をランダムに生成することができたり，あるいは，あるNP問題に対する難しい正例題のみをランダムに生成することができれば，暗号システムにおける暗号鍵の生成手法への応用も考えられる．

ひとくちに「ランダムに生成する」といっても，どのような例題をどのような分布で生成するかによって，例題生成手法の良し悪しが決まってくる．究極的には，与えられた確率分布に対して，それに従った分布で例題を生成できることが望ましく，そのための一般的な研究も行われている[1),2)]．また一方で，特徴的な分布に従って例題を生成する手法や，ある種の性質を持った問題だけを生成する手法の研究も重要である．たとえば，解が1つに定まるような例題のみを生成することもしばしば重要になってくる．というのも，複数の解がある例題は，暗号鍵の生成手法への応用を考えた場合，セキュリティ上に問題が生じる可能性があるからである．また，3SAT問題においては，解が1つに

† 株式会社増進会出版社
†† 東京工業大学情報理工学研究科計算工学専攻
††† 東京工業大学情報理工学研究科数理・計算科学専攻

定まるような例題に対し，その解を否定するような項を1つ付け加えれば負の例題になる[3]．したがって，負例題の生成手法としても利用できる可能性が出てくるのである．本論文では，単純な生成アルゴリズムGを考え，ある条件のもとでは，そのアルゴリズムが3SATの正例題のうち，解を1つしか持たないものを生成する確率が非常に高いことを示す．また，その事実から，3SATの正例題をほぼ等確率で出力することも導く．この性質は，その条件を少し緩めた場合でも成り立つことが予想されるが，実験によりその可能性を示す．

具体的には，次のような単純な生成アルゴリズムGを考える．

アルゴリズム G

input: 変数の数 n，項の数 m;
begin
　真偽値割当 $\mathtt{t} \in \{0,1\}^{\mathtt{n}}$ を等確率で生成;
　t に対して真となる項を m 項，等確率で生成;
　生成した m 項を ∧ で連結した式 f を作り，
　f と t を出力する;
end.

与えられた n, m に対し，このアルゴリズムGの出力を $\mathtt{G}(n,m)$ と記述し，また，出力のうち論理式fの方を $\mathtt{G}(n,m)_{\mathtt{f}}$，充足割当t（すなわちfの解）の方を $\mathtt{G}(n,m)_{\mathtt{t}}$，とそれぞれ表すことにする．これらはGの計算中に利用した乱数に依存して決まるランダム変数である．

アルゴリズムの作り方から明らかなように，すべての n 変数 m 項の充足可能論理式 f が，Gにより生成可能である．つまり，$\mathtt{G}(n,m)_{\mathtt{f}} = f$ となる確率は0ではない．しかも，すべての真偽値割当 $t \in \{0,1\}^n$ と，その t を解に持つすべての f に対し，$\mathtt{G}(n,m) = (f,t)$ となる確率は等しい（その確率を $p_{n,m}$ とする）．ただし，f が生成される確率，つまり $\mathtt{G}(n,m)_{\mathtt{f}} = f$ となる確率を考えると，それは f の充足解の個数 k に依存し，生成確率は $kp_{n,m}$ となる．したがって，Gはすべての充足可能論理式を等確率で生成するとはいえない．

ところが，項の数 m が，変数の数 n に比べて大きい場合，充足解の個数は一般には少なくなることが容易に予想できる．もし仮に，ほとんどの論理式が1つしか解を持たないとしたら，アルゴリズムGは，ほとんどの場合，充足可能論理式を等確率 $p_{n,m}$ で出力することになる．本論文では，m が n に対してどの程度大きくなると，ほとんどの充足可能論理式が1つしか解を持たなくなるかを解析した．

まず，理論的に，入力が $m \geq d_1 n^2$（ただし d_1 はある正定数）のとき，Gは充足解が1つしかない論理式を高い確率で生成することを証明した．この性質は，項数 m に関する条件を少し緩めても成り立つように思われる．そこで，実際にランダムに論理式を生成し，充足解が1つしかない論理式を高い確率で生成する場合の m と n の関係を調べた．その結果，$m \geq d_2 n \log n$（ただし d_2 はある正定数）のとき，解が1つしかない例題が高い確率で生成されることを示した．

2. 準　　備

ここでは，基本的な用語や記号の定義を行う．

次のような形の論理式を，3和積標準形（3 Conjunctive Normal Form, 3CNF）論理式という．

$$(x_2 \vee \overline{x_4} \vee x_1) \wedge (\overline{x_3} \vee x_1 \vee x_2) \wedge (x_1 \vee x_2 \vee \overline{x_4}).$$

ここで，$x_1, \ldots$ を変数，x_i またはその否定 $\overline{x_i}$ をリテラル，$(x_2 \vee \overline{x_4} \vee x_1)$ のように3つのリテラルの∨からなる式を項という．ただし本論文では，各項における変数の重複を許さず，また変数の順序を入れ換えても同じ項と見なすことにする．たとえば，上の例の1番目と3番目の項は同じと見なされる．したがって，n 変数の3CNFにおける項は，$8\binom{n}{3}$ 種類あることになる（一方，与えられた割当 t に対し，その割当で充足される3CNF項は $7\binom{n}{3}$ 通り．この数を以下では N_0 と表す）．また項に関しても，重複を許さず，項の順序を入れかえた式は同じものと考えることにする．したがって，アルゴリズムGにおいて，「等確率で生成」というのは，正確には，m 個の項を集合として見たうえで等確率で生成することを意味している．

n 変数の論理式 f に対して，長さ n の0,1-列 $t \in \{0,1\}^n$ を，f に対する（真偽値）割当という．割当 t に対する $f(t)$ の値は，f の各変数 x_i に対して，t の i ビット目（t_i）を代入して得られる値である．もし $f(t) = 1$ となるとき，t を f の充足割当（あるいは充足解）という．さらに，充足割当を持つ論理式を充足可能論理式という．3SATは充足可能3CNF論理式の集合で，3SAT問題とは，与えられた3CNF論理式が3SATに入るかどうか（すなわち充足可能かどうか）をテストする問題である．それに対し，与えられた3CNF論理式に対して，もし存在するならば充足割当を求める問題を，3SAT探索問題と呼ぶ．

KoutsoupiasとPapadimitoriou[4]は，3SAT探索問題に対する次のような単純なグリーディ・アルゴリズムAを解析した．

アルゴリズム A

input: n 変数 m 項の 3CNF f;
begin
　真偽値割当 $\mathtt{t} \in \{0,1\}^{\mathtt{n}}$ を等確率で生成;
　while $\exists \mathtt{t}' \in N(\mathtt{t})$ [$\#\mathtt{c}(\mathtt{f},\mathtt{t}') > \#\mathtt{c}(\mathtt{f},\mathtt{t})$]
　　do { $\mathtt{t} := \mathtt{t}'$; }
　$\mathtt{f}(\mathtt{t}') = 1$ ならば $\mathtt{t}'$ を出力;
　（そうでない場合には「解なし」と答える）
end.

ただしここで，$\#\mathtt{c}(f,t)$ は，f 中の項で t によって真となる項の数とする．また $N(t)$ は t と 1 ビットだけ異なる割当の集合を表す．この記法は，これ以降の議論でも用いられる．

本論文ではこの充足解探索アルゴリズム A の性質を利用し，単純な例題生成アルゴリズム G の解析を行う．これはランダム・アルゴリズムなので，その計算は使用した乱数値に依存する．以下では，G は一様分布乱数を用いていると仮定し，$\Pr_{\mathtt{G}}\{\cdots\}$ で，その場合に事象「…」が成立する確率を表す．

3. 理論的解析

各 $n, m \geq 1$ に対し，$F_{n,m}$ を充足可能な n 変数，m 項の 3CNF の全体とする．また $F_{1,n,m}$ を $F_{n,m}$ の中で充足解が 1 つしかない 3CNF の全体とする．この章では，次の定理を証明する．

定理 3.1　十分大きい n に対して，$m \geq d_1 n^2$ のとき，

$$\Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} \in F_{1,n,m}\} > 1 - 2^{-n}.$$

補注. d_1 は定数．かなり大雑把だが，たとえば $n \geq 6$, $d_1 = 6000$ ならば，定理は成り立つ．

ここでは，この定理の証明を与える．以下では，定理の条件を満たす n, m を任意に固定したうえで議論する．そこで，以下の議論では，とくに断らなくても，n は十分大きく，$m \geq d_1 n^2$ が成り立っているものと仮定する．また，$F_{n,m}$, $F_{1,n,m}$ 添字の n, m も省略して，簡単に F, F_1 と記述することにする．

G を直接評価するのは難しいので，次のような例題生成アルゴリズム $\mathtt{G}_0$ を考える．この $\mathtt{G}_0$ が m 項の式を出力する（この事象を MC と表す）とき，G と等価な評価ができることを示し，この $\mathtt{G}_0$ について理論的解析を行う．

アルゴリズム $\mathtt{G}_0$

input: 変数の数 n，項を選ぶ確率 p；
begin
　真偽値割当 $\mathtt{t} \in \{0,1\}^{\mathtt{n}}$ を等確率で生成；
　for each c: c は t で真となる項 **do**
　　{ 確率 p で $\mathtt{f} \leftarrow \mathtt{f} \wedge \mathtt{c}$,
　　確率 $1-\mathtt{p}$ で $\mathtt{f} \leftarrow \mathtt{f}$; }
　f と t を出力;
end.

補題 3.2　F の任意の要素 f に対して

$$\forall p \Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} = f\} = \Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} = f \,|\, \mathrm{MC}\}.$$

証明.　G と $\mathtt{G}_0$ がともに真理値割当 t を選んだとする．このとき，割当 t で充足する任意の m 項の式 f_t に対し，$\mathtt{G}_0$ と G が f_t を出力する確率を考える．

$$\begin{aligned}
&\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} = f_t \,|\, \mathrm{MC} \wedge \mathtt{G}_0(n,p)_{\mathtt{t}} = t\} \\
&= \frac{\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} = f_t \wedge \mathtt{G}_0(n,p)_{\mathtt{t}} = t \wedge \mathrm{MC}\}}{\Pr_{\mathtt{G}_0}\{\mathrm{MC}\}\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{t}} = t\}} \\
&= \frac{\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} = f_t \wedge \mathtt{G}_0(n,p)_{\mathtt{t}} = t\}}{\Pr_{\mathtt{G}_0}\{\mathrm{MC}\}\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{t}} = t\}} \\
&= \frac{p^m(1-p)^{N_0-m}}{\binom{N_0}{m}p^m(1-p)^{N_0-m}} = \frac{1}{\binom{N_0}{m}} \\
&= \Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} = f_t \,|\, \mathtt{G}(n,m)_{\mathtt{t}} = t\}.
\end{aligned}$$

つまり G と $\mathtt{G}_0$ でそれぞれ真理値割当 t が選ばれたとき，G で f_t が出力される確率と，$\mathtt{G}_0$ で出力された式の項数が m になったとき式が f_t である確率は等しい．そこで任意の $f \in F$ について f の充足解の集合を $T(f)$ と表すと，

$$\begin{aligned}
&\Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} = f\} \\
&= \sum_{t \in T(f)} \Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} = f \,|\, \mathtt{G}(n,m)_{\mathtt{t}} = t\}, \\
&\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} = f \,|\, \mathrm{MC}\} \\
&= \sum_{t \in T(f)} \Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} = f_t \,| \\
&\qquad \mathrm{MC} \wedge \mathtt{G}_0(n,p)_{\mathtt{t}} = t\}
\end{aligned}$$

となり，$\Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} = f\} = \Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} = f|\mathrm{MC}\}$ がいえる．□

この補題から，G の評価を，$\mathtt{G}_0$ で m 項の式が出力されたときと置き換えて議論することができる．補題 3.2 は任意の p について成り立つので，以下では $p = m/N_0$ とする．このとき事象 MC が起きる確率 $\Pr_{\mathtt{G}_0}\{\mathrm{MC}\}$ の解析は，次のようになる．

$$\begin{aligned}
&\Pr_{\mathtt{G}_0}\{\mathrm{MC}\} \\
&= \binom{N_0}{m}p^m(1-p)^{N_0-m} \\
&= \frac{N_0!}{m!(N_0-m)!}\left(\frac{m}{N_0}\right)^m\left(\frac{N_0-m}{N_0}\right)^{N_0-m} \\
&\approx \frac{\sqrt{2\pi}N_0^{N_0+1/2}e^{-N_0}}{\sqrt{2\pi}m^{m+1/2}e^{-m}}
\end{aligned}$$

$$\cdot \frac{1}{\sqrt{2\pi}(N_0-m)^{N_0-m+1/2}e^{-(N_0-m)}} \cdot \frac{m^m}{N_0^m} \cdot \frac{(N_0-m)^{N_0-m}}{N_0^{N_0-m}} = \sqrt{\frac{N_0}{2\pi m(N_0-m)}} \geq \frac{1}{\sqrt{2\pi m}}.$$

アルゴリズム A の解析において，Koutsoupias と Papadimitoriou[4]は，A がうまく働くための十分条件を考え，ほとんどの $f \in F$ が，その条件を満たしていることを示した．それによって，ほとんどの $f \in F$ に対して，A がうまく充足割当を見つけることを証明したのである．ここでも，その性質を利用して定理を証明する．

まず，用語や記号を少し用意する．任意の割当 $t, t' \in \{0,1\}^n$ に対し，t と t' のハミング距離（値の食い違うビット数）を $\mathrm{d}(t,t')$ と表す．任意の ε と，割当 t 対して，$\mathrm{d}(t,t') \leq (1/2+\varepsilon)n$ となる割当 t' を，t に対して ε-good な割当という（なお，以下で具体例が必要な場合には，$\varepsilon = 1/6$ を考える）．さらに $N_\varepsilon(t)$ を，t に対する ε-good な割当の全体とする．

いま，任意の $f \in F$ を考え，t を f の充足割当（の任意の1つ）とする．この f, t が次の条件を満たすとき，f は $N_\varepsilon(t)$ において単調収束構造を持つということにする．

$$\forall t_2 \in N_\varepsilon(t),\ \forall t_1 \in N(t_1) \left[\begin{array}{l} \mathrm{d}(t_2,t) < \mathrm{d}(t_1,t) \\ \quad \Rightarrow\ \#\mathrm{c}(f,t_2) > \#\mathrm{c}(f,t_1) \end{array} \right] \qquad (1)$$

さらに，f に対して，この条件を満たすような充足割当 t が存在するとき，ただ単に，f は単調収束構造を持つ，ということにする．

$N_\varepsilon(t)$ において単調収束構造を持つ論理式の集合 $F(\to t)$ と，そうでない論理式の集合 $F(\not\to t)$ に分けて考える．この $F(\not\to t)$ に関する次の補題が，ここでの理論的解析で要となってくる．

補題 3.3

$$\mathrm{Pr}_{\mathrm{G}_0}\{\mathrm{G}_0(n,p)_{\mathtt{f}} \in F(\not\to t) \,|\, \mathrm{MC}\} < \frac{\sqrt{2\pi m}\cdot n2^n}{e^{cmn^2/N_0}}.$$

補注. この確率は「t が出力される」という条件付き確率である．なお，c は ε による決まる定数で，

$$c \approx \left(1-\sqrt{1-(1/2-\varepsilon)^2}\right)^2/6.$$

証明. この補題は，文献4)とほとんど同じ論法で示すことができる．考え方としては，アルゴリズム G_0 の最初の段階で，$\mathtt{t}$ に t が選ばれたと仮定し，この場合に m 項の式が出力され，$\mathrm{G}_0(n,p)_{\mathtt{f}} \in F(\not\to t)$ となる確率を求めればよい．つまり，生成された f_t が，ある $t_1 \in N_\varepsilon(t) - \{t\}$ と，t_1 より1ビットだけ t に近い割当 t_2 の組に対し，式(1)を満たさない確率である．

いま，$N_\varepsilon(t) - \{t\}$ の任意の要素 t_1 と，t_1 より1ビットだけ t に近い割当 t_2 を考え，生成された f_t が，この t_1, t_2 に対して式(1)を満たさない確率を評価してみよう．以下では，この確率を $\mathrm{Pr}_{\mathrm{G}_0}\{\text{no gain}\}$ と表すことにする．なお，t_1 は，t と k ビットで一致しているものとする（ただし $k \geq (1/2-\varepsilon)n$）．

解析のために，まず次のような項の集合 $C_\oplus, C_\ominus$ を定義する．

$$C_\oplus = \{c \,|\, c(t)=1, c(t_1)=0, c(t_2)=1\},$$
$$C_\ominus = \{c \,|\, c(t)=1, c(t_1)=1, c(t_2)=0\}.$$

この $C_\oplus, C_\ominus$ の要素数は，それぞれ $\binom{n-1}{2}$, $\binom{n-1}{2} - \binom{k}{2}$ だが[4]，前者を $N_\oplus$，後者を $N_\ominus$ とする．

また，G_0 の実行で i 番目に生成された項を c_i とし，各 i $(1 \leq i \leq N_0)$ に対して次の確率変数を定義する．

$$X_i = \begin{cases} 1, & c_i \in C_\oplus \text{ のとき}, \\ 0, & \text{その他のとき}. \end{cases}$$

$$Y_i = \begin{cases} 1, & c_i \in C_\ominus \text{ のとき}, \\ 0, & \text{その他のとき}. \end{cases}$$

さらに，$X = \sum_{i=1}^m X_i$, $Y = \sum_{i=1}^m Y_i$ と定義すると，事象 no gain が起きる確率を，$\mathrm{Pr}_{\mathrm{G}_0}\{\text{no gain}\} = \mathrm{Pr}_{\mathrm{G}_0}\{Y \leq X\}$ と表すことができる．

しかも，任意の M に対して，次の式が成り立つ．

$$\mathrm{Pr}_{\mathrm{G}_0}\{Y \leq X\} \leq \mathrm{Pr}_{\mathrm{G}_0}\{Y \leq M \vee M \leq X\} \leq \mathrm{Pr}_{\mathrm{G}_0}\{Y \leq M\} + \mathrm{Pr}_{\mathrm{G}_0}\{M \leq X\} \qquad (2)$$

一方，各項は確率 p で式に加えるので，f_t に含まれる $C_\oplus, C_\ominus$ の要素の個数の分布は，二項分布となる．したがって，Chernoff bound を使えば，式(2)は次のように評価できる[4]．

$$\mathrm{Pr}_{\mathrm{G}_0}\{Y \leq M\} + \mathrm{Pr}_{\mathrm{G}_0}\{M \leq X\} \leq e^{-(M-pN_\ominus)^2/3pN_\ominus} + e^{-(M-pN_\oplus)^2/3pN_\oplus} \qquad (3)$$

ここで，$M = m\sqrt{N_\ominus N_\oplus}/N_0$ と選ぶことによって，上式の右辺の各項が等しくなり，さらに各々を e^{-cmn^2/N_0} で抑えることができるので（付録参照），結局，$\mathrm{Pr}_{\mathrm{G}_0}\{\text{no gain}\} \leq 2e^{-cmn^2/N_0}$ が得られる．

また，

$$\mathrm{Pr}_{\mathrm{G}_0}\{\text{no gain} \,|\, \mathrm{MC}\} = \frac{\mathrm{Pr}_{\mathrm{G}_0}\{\text{no gain} \wedge \mathrm{MC}\}}{\mathrm{Pr}_{\mathrm{G}_0}\{\mathrm{MC}\}} \leq \frac{2\sqrt{2\pi m}}{e^{cmn^2/N_0}}.$$

最後に，t_1 と t_2 の組が，たかだか $n2^{n-1}$ 通りしかないことを使えば，目標の評価式を次のように導き出すことができる．

$$\begin{aligned}&\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F(\not\to t)\,|\,\mathrm{MC}\}\\ &= \Pr_{\mathtt{G}_0}\{\exists t_1,t_2\,[\text{no gain}]\,|\,\mathrm{MC}\}\\ &\le (2\sqrt{2\pi m}\cdot e^{-cmn^2/N_0})\cdot n2^{n-1}\\ &= \sqrt{2\pi m}\cdot n2^n e^{-cmn^2/N_0}.\end{aligned}$$

□

次の関係は，この補題から簡単な計算で導かれる．

系 3.4

$$\begin{aligned}&\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F(\not\to t)\,|\,\mathrm{MC}\}\\ &< \frac{1}{2^{\alpha n}}\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F(\to t)\,|\,\mathrm{MC}\}.\end{aligned}$$

補注. ただし，ここでは $\alpha = (6cd_1/7)\log e - 3$ とする（実際はもう少し小さい値でもよい）．たとえば，$\varepsilon = 1/6$ と決めると，$c \approx (1-2\sqrt{2}/3)^2/6 \approx 0.000545$ となるので，$d_1 = 6000$ ととれば，十分 $\alpha \ge 1$ となる．

証明. $m > d_1 n^2$ なので $\frac{m}{N_0} > \frac{6d_1}{7n}$．また，$m \le N_0 = 7\binom{n}{3} < \beta n^3$ なので，適当な β' を用いて，$\sqrt{2\pi m} < \beta' n^{\frac{3}{2}}$ とできる．よって

$$\begin{aligned}&\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F(\not\to t)\,|\,\mathrm{MC}\}\\ &= \sqrt{2\pi m}\cdot n2^n e^{-cmn^2/N_0}\\ &< \beta' n^{\frac{3}{2}}\cdot n2^n e^{-6cd_1 n/7}\\ &< 2^n\left(\frac{2}{e^{6cd_1/7}}\right)^n < 2^{-\gamma n}\end{aligned}$$

となる（ただし $\gamma = (6cd_1/7)\log e - 2$）．一方，$\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F(\to t)\,|\,\mathrm{MC}\} \ge 1 - 2^{-\gamma n}$ なので，$\alpha = (6cd_1/7)\log e - 3$ とすれば目標の式が成り立つ．□

定理 3.1 の証明. まず，次の 2 つの事実を示す．

事実 1 $F - F_1 \subseteq \cup_{t\in\{0,1\}^n} F(\not\to t)$.

証明. $F - F_1$ の任意の要素，すなわち，2 つ以上の充足割当を持つ $f \in F$ を考え，t_1, t_2 を f の異なる充足割当とする．ここで，$\varepsilon = 1/6$，$n \ge 6$ と仮定すると，$N_\varepsilon(t_1) \cap N_\varepsilon(t_2) \ne \emptyset$．したがって，どちらかの $N_\varepsilon(t_i)$ においては，f は単調収束構造を持たない．ゆえに，$f \in F(\not\to t_i)$．よって $F - F_1 \subseteq \cup_{t\in\{0,1\}^n} F(\not\to t)$.

□

事実 2 任意の異なる割当 t, t' に対し，$F(\to t) \cap F(\to t') = \emptyset$.

証明. もし $f \in F(\to t) \cap F(\to t')$ とすると，f は $N_\varepsilon(t)$ においても，$N_\varepsilon(t')$ においても単調収束構造を持たなければならない．しかし，$\varepsilon = 1/6$，$n \ge 6$ と仮定すると，$N_\varepsilon(t) \cap N_\varepsilon(t') \ne \emptyset$．よって，$f$ は $N_\varepsilon(t)$ と $N_\varepsilon(t')$ の両者において，単調収束構造を持つことはできない．□

以上の事実，ならびに系 3.4 を使えば，確率 $\Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} \notin F_1\}$ に対する上限を，次のように導き出すことができる（ただし，$\varepsilon = 1/6$，$n \ge 6$，$m \ge 6000n^2$ と仮定）．

$$\begin{aligned}&\Pr_{\mathtt{G}}\{\mathtt{G}(n,m)_{\mathtt{f}} \notin F_1\}\\ &= \Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \notin F_1\,|\mathrm{MC}\}\\ &= \Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F - F_1\,|\mathrm{MC}\}\\ &\le \Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in \cup_{t\in\{0,1\}^n} F(\not\to t)\,|\mathrm{MC}\}\\ &\le \sum_{t\in\{0,1\}^n} \Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F(\not\to t)\,|\mathrm{MC}\}\\ &\le \sum_{t\in\{0,1\}^n} \frac{1}{2^n}\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F(\to t)\,|\mathrm{MC}\}\\ &\le \frac{1}{2^n}\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in \cup_{t\in\{0,1\}^n} F(\to t)\,|\mathrm{MC}\}\\ &\le \frac{1}{2^n}\Pr_{\mathtt{G}_0}\{\mathtt{G}_0(n,p)_{\mathtt{f}} \in F\,|\mathrm{MC}\} = \frac{1}{2^n}.\end{aligned}$$

定理の式は，これより明らか．□

アルゴリズム G は，割当 t ごとに，その割当を充足割当として持つ 3CNF を等確率で生成する．この事実を用いると，定理 3.1 は，次のようにもいいかえられる．

系 3.5 十分大きい n に対して，$m \ge d_1 n^2$ のとき，$\|F_1\| \ge (1-2^{-n})\|F\|$.（ただし，$\|F_1\|, \|F\|$ は，各々，F_1, F の要素数を表す）．

4. 実験による解析

実際にアルゴリズム G を用いて，n 変数 m 項の論理式を生成し，高い確率で解が 1 つしかない式が生成される項の数を調べる実験を行った．その結果について報告する．

実験方法

アルゴリズム G で生成された 3CNF 式が，複数の解を持つかを調べる実験を行った．ただし，解の探索を全解探索で行うことは不可能なので，グリーディ・アルゴリズム A を使った近似探索を行った．具体的には，生成された式 f（その充足割当を t とする）を A に与え，ランダムに選んだ k 個の初期割当から出発し，t 以外の解が出力されるかを確かめたのである．その場合の k として，どのような値が妥当であるかを，まず次のような予備実験を行って求めた．

予備実験では，2 つの解を持つ 3CNF 式を人工的に作成し，その式に対して，上記の近似探索で 2 つ以上の充足割当を確認するため必要な，A の実行回数 k を調べた．なお，この実験では簡単のために解の 1 つを

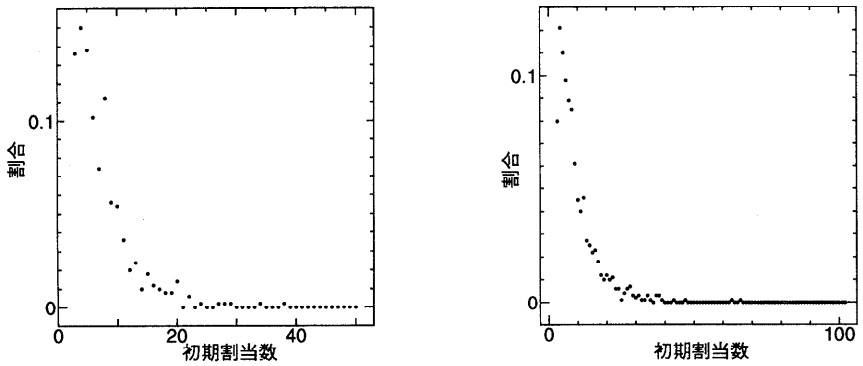


図1 複数解の存在確認に k 回の実行が必要な式の割合（左： $n=50$, $m=600$, 右：$n=90$, $m=1080$）
Fig. 1 Ratio of formulas that requires k times greedy algorithm to find two solutions (left: $n=50$, $m=600$, right: $n=90$, $m=1080$).

1^n に固定し，もう片方の解を $\{0,1\}^n-\{1^n\}$ からランダムに選んだ．

すると，n にあまり関係なく m/n がある程度（具体的には10程度）以上大きければ，k の増加とともに k 回の実行が必要な式の割合が急速に減少することが分かった（図1のグラフ参照）．具体的には，今回の実験での変数・項の数の範囲では20回程度行えば，ほぼ95%以上の確率で，異なる充足割当を確認できることが確認できた．そこで，グリーディ・アルゴリズム A を20回適用し，複数解の存在を確かめることにした．

理論的な解析からも分かるように，ある割当 t を固定し，その割当で充足させられる式に対して，複数解が存在しない確率を調べればよい．そこで，今回の実験では，単一の解となる割当を $t=1^n$ と固定し，t に充足される項を重複しないように m 個選んで生成した論理式が t 以外の充足割当を持たない確率を調べた．

なお，実験はCRAY C916上で行い，乱数発生にはC916上のCray Standard Cの組込み関数rand()を用いた．

実験結果について

実験では，変数の数 n，項数 m の各値に対して，1000個の式をランダムに生成し，単一解を持つ式（充足割当を1つしか持たない式）が生成された回数 p を調べた．したがって，$r=p/1000$ が，単一解を持つ式が生成された割合である．ただし，変数の数 n は10から90まで10刻みに調べた．その中で $n=90$ のときの項変数比 m/n と r の関係を図2に示す．この図から分かるように，項変数比 m/n の増加にともない緩やかに1に近づく．

今回の実験では，使用したアルゴリズムがランダム

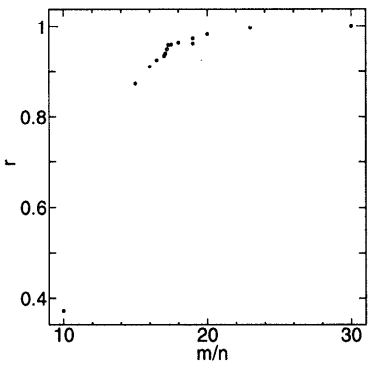


図2 $n=90$ のときの r の変化
Fig. 2 Relation between r and m/n on $n=90$.

表1 単一解を持つ式が生成された回数 p
Table 1 Number of formulas with unique solution.

n							
10	m	113	115				
	p	953	950				
20	m	270	272	275	276	280	
	p	947	956	958	963	952	
30	m	420	429	435	441	444	447
	p	945	936	944	936	947	953
40	m	600	604	608	610	612	616
	p	945	945	941	950	962	948
50	m	790					
	p	954					
60	m	990					
	p	951					
70	m	1155	1162	1169	1172	1176	1183
	p	947	947	944	950	943	951
80	m	1376	1384				
	p	952	952				
90	m	1548					
	p	950					

アルゴリズムなので，単一解を持つ式（解を1つしか持たない3CNF）が生成される割合 r は，決まった値にはならない．そこで，変数の数 n に対し，r の値が小数点以下第3桁を四捨五入して0.95となる項数 m

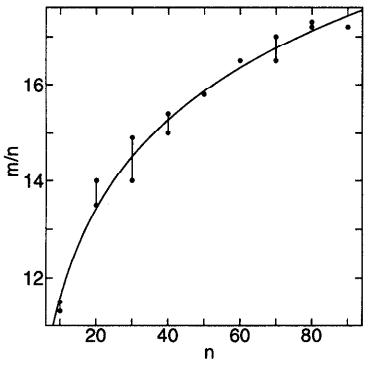


図3 単一解を持つ式の割合が 0.95 になる場合の n と m/n の関係
Fig. 3 relation between n and m/n where ratio of number of fomulas which has only one solution.

の範囲を求めることにした．**表1** に，その m の範囲と，実験により実際に得られた p（単一解を持つ式が生成された回数）の値を示す．

このようにして得られた m の上限，下限を用い，横軸に変数の数 n，縦軸に項変数比 m/n としてグラフにプロットすると，**図3** のようになる．

このグラフから項変数比 m/n が $\log n$ に比例しているように見える．そこで最小自乗法を用いて近似をとると，$m/n = 5.46 + 2.66 \ln n$ という結果が得られた．つまり，この実験は，$m \geq 5.46n + 2.66n \ln n$ であれば，単一解を持つ式が出力される確率が 0.95 以上になることを示している．

5. おわりに

3充足可能性判定問題3SATの正例題生成を行う単純なアルゴリズムを解析し，項数 m が変数 n に対して $m \geq d_1 n^2$ の例題を生成する場合には，ほとんどの場合，充足解が1つしかない論理式が生成されることを証明した．また，条件を弱めて $m \geq d_2 n \log n$ の場合でも，充足解が1つしかない論理式が生成されるという実験的な結果を示した．今後の課題としては，この実験結果を理論的に証明することが重要だろう．

最初にも述べたように，充足解が1つしかない論理式の生成には，負の例題の生成という重要な応用がある．一般に負例題の生成は，正例題の生成よりもさらに難しいといわれている．たとえば，NP $\neq$ co-NP の仮定のもとでは，3SAT 問題に対する負例題をすべて生成する多項式アルゴリズムは存在しない（それに対し，今回の単純な生成アルゴリズムは，生成確率のばらつきはあるにせよ，正例題をすべて生成している）．また，今までにもいろいろ研究されているが（たとえば，文献5) 参照），決め手になる手法は見つかっていない．充足解が1つしかない論理式を生成し，その充足解を殺すような項をいくつか追加して負例題を作る手法は，一般的で強力な負例題の生成手法として使用できるものと考えられる．

残念ながら，本論文で理論的に示せる条件（$m \geq d_1 n^2$）のもとでは，定理の証明から明らかなように，グリーディ・アルゴリズムで解けるような例題が数多く生成されることになる．したがって，それから生成された正例題，それに負例題も，比較的簡単に解かれてしまうだろう．しかし，実験結果は条件を緩められる可能性を示しており，その場合には，グリーディ・アルゴリズムでは難しい式が多く生成されるかもしれない．こうした点の理論的解析が，これからの重要な課題である．

謝辞 本論文の初期の版における誤りや不備な点を指摘してくださいました査読者の方々に感謝いたします．

参考文献

1) Karg, C., Köbler, J. and Schuler, R.: The complexity of generating test instances, *Proc. STACS'97*, Lecture Notes in Computer Science (1997). to appear.
2) Watanabe, O.: Test instance generation for promised NP search problems, *Proc. 9th Structure in Complexity Theory Conference*, New York, pp.205–216, IEEE (1994).
3) 元木光雄：充足不可能な論理式例題の生成に関する研究，東京工業大学，卒業論文，(1996).
4) Koutsoupias, E. and Papadimitriou, C.: On the greedy algorithm for satisfiability, *Infom. Process. Lett.*, Vol.43, pp.53–55 (1992).
5) 高木和哉，岩間一雄：リテラルの出現回数を制限した充足不能な 3CNF 式，電気情報通信学会技術研究報告 (1995).

付　録

式 (3) の右辺を上から抑える．本文中でも説明したように，$M = m\sqrt{N_\ominus N_\oplus}/N_0$ と選ぶことによって，式 (3) の右辺の各項は，$c^{-(p/3)(\sqrt{N_\oplus}-\sqrt{N_\ominus})^2}$ と等しくなる．

一方，

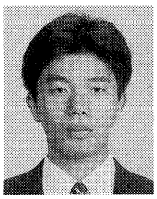
$$
\begin{aligned}
& e^{-\frac{p}{3}\left(\sqrt{N_\oplus}-\sqrt{N_\ominus}\right)^2} \\
&= e^{-\frac{p}{3}\left(\sqrt{\binom{n}{2}}-\sqrt{\binom{n}{2}-\binom{k}{2}}\right)^2} \\
&= e^{-\frac{p}{6}\left(\sqrt{(n-1)(n-2)}-\sqrt{(n-1)(n-2)-k(k-1)}\right)^2}
\end{aligned}
$$

$$\leq e^{-\frac{p}{6}\left(\sqrt{(n-1)(n-2)}-\sqrt{(n-1)(n-2)-n'(n'-1)}\right)^2}$$

となる（ただし，$n'=n(1/2-\varepsilon)$）．ここで，十分大きな n を考えると，$n-1, n-2$ を n, $n'-1$ を n' で近似しても構わないので，以下のように目的の上界が導き出せる．

$$\begin{aligned} & e^{-\frac{p}{6}\left(\sqrt{(n-1)(n-2)}-\sqrt{(n-1)(n-2)-n'(n'-1)}\right)^2} \\ \approx\ & e^{-\frac{p}{6}\left(\sqrt{n^2}-\sqrt{n^2-(1/2-\varepsilon)^2n^2}\right)^2} \\ =\ & e^{-\frac{p}{6}n^2\left(1-\sqrt{1-(1/2-\varepsilon)^2}\right)^2} \\ =\ & e^{-cpn^2}. \end{aligned}$$

(平成 9 年 4 月 2 日受付)
(平成 10 年 3 月 6 日採録)



望月　厚

1970 年生．1992 年東京工業大学工学部情報工学科卒業．1994 年同大学大学院理工学研究科情報工学専攻修士課程修了．同年，(株) 増進会出版社に入社．

元木　光雄

1970 年生．1996 年東京工業大学工学部電気電子工学科卒業．1998 年同大学大学院情報理工学研究科計算工学専攻修士課程修了．同年，同大学大学院情報理工学研究科数理・計算科学専攻博士後期課程進学．

渡辺　治（正会員）

1958 年生．1980 年東京工業大学理学部情報科学科卒業．1982 年同大学大学院理工学研究科情報科学専攻修士課程修了．同年，同博士課程中退．工学博士．1982 年東京工業大学理学部助手．1990 年同工学部助教授．1997 年東京工業大学大学院情報理工学研究科教授．専門は計算の複雑さの理論とその応用．電気情報通信学会，EATSC，LA 各会員．