3 充足可能性判定問題 3SAT の単一解を持つ正例題生成手法の解析

元木 光雄

東京工業大学 情報理工学研究科 計算工学専攻

〒 152-8552 東京都目黒区大岡山 2-12-1

mmotoki@cs.titech.ac.jp

上原 隆平

東京女子大学 情報処理センター

〒 167 東京都杉並区善福寺 2-6-1

uehara@twcu.ac.jp

あらまし　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問題— 3SAT の正例題生成問題 — を考える．本論文では，単純な例題生成アルゴリズムによって n 変数，m 項の 3SAT の要素を生成したとき，単一解を持つ例題が生成されるための項数 m の条件を調べた．さらに，項数を減らすためのアルゴリズムの改良を行い，項数の条件を調べた．

キーワード　充足可能性判定問題，正例題生成手法，単一解を持つ例題，理論的解析と実験的解析

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

Mitsuo Motoki

Dept. of Computer Science

Tokyo Institute of Technology

Meguro-ku Ookayama 2-12-1, Tokyo 152, Japan

mmotoki@cs.titech.ac.jp

Ryuhei Uehara

Tokyo Woman's Christian University

Center for Information Science

Zempukuji 2-6-1, Suginami-ku, Tokyo 167, Japan

uehara@twcu.ac.jp

Abstract. For the 3-Satisfiability Problem (3SAT), we consider a problem of generating its positive instance (and its solution) randomly. For the problem, it is desirable 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.

Key Words 3SAT, Positive instance generation algorithm, unique solution instance, theoretical analysis and an experimental result

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取り上げ，その解析を行う．NP 問題の正例題生成手法とは NP 集合 X に対し，X の元 x と $x \in X$ の証拠（普通「x の解」と呼ばれる）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 の場合は，集合 3SAT（充足可能な 3CNF 論理式の集合）に属する 3CNF 論理式 f と，その f の充足割当 t をランダムに生成する問題がそれにあたる．また，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，例題生成問題はその元の問題を解くアルゴリズムのテストや性能評価に用いられる．さらに，NP 問題に対するその他の様々な性質を実験的に調べる場合にも例題生成手法が有効であると思われる．また，暗号システムにおける暗号鍵の生成への応用も考えられる．すなわち，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，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，これを暗号鍵に使用することで解読することが難しい暗号ができると考えられる．

「ランダムに生成する」と言っても，どのような性質の例題をどのような分布で生成するかによって，例題生成手法の善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い．その一方で，特徴的な分布にしたがって例題を生成する手法や，ある種の性質を持った例題のみを生成する手法も求められることがある．本論文では，こうした具体的な分布の一例として，単一解を持つ 3CNF 論理式を生成する問題を議論する．単一解を持つ例題はそれ自身は正例題であるが，3SAT 問題の場合，その解を否定する項を追加することによって，負の例題生成に用いることができる [1]．一般に NP 問題の負の例題をもれなく生成することは不可能と予想されているので，このような手法は負の例題生成手法として有効な手法の一つといえよう．

充足可能な 3CNF 論理式とその充足解 $\hat{t}$ を生成する手法として，まず次の様な単純な生成アルゴリズム G を考えた．

アルゴリズム G

input: 変数の数 n, 項の数 m;

begin

　真偽値割当 $\hat{t} \in \{0,1\}^n$ を等確率で生成;

　for $i = 1$ **to** m

　　do $\hat{t}$ で充足される項を等確率で生成し，

　　f の i 番目の項として追加;

　$(f, \hat{t})$ を出力;

end.

アルゴリズムの作り方から，全ての n 変数 m 項の充足可能な 3CNF 論理式が G によって出力可能なことが明らかである．また，一般に項の数 m が増えると充足解の数は少なくなることが予想されるので，十分大きな m に対しては G は単一解を持つ論理式を生成する．問題は，m がどの程度大きければよいかという点である．望月・元木・渡辺 [2] は $m = \Omega(n^2)$ ならば高い確率で単一解を持つ式を出力することを理論的に示した．さらにこの文献では，実験的な解析により $m = \Omega(n \ln n)$ でも十分高い確率で単一解を持つ式を出力することが示されている．本論文では，まずこの実験的に得られた結果を理論的に証明した．まず，理論的に $m = \Omega(n \ln n)$ のときに高い確率で単一解を持つ式が出力されることを示した．さらに，$m = \Omega(n \ln n)$ が必要であることも証明した．さらに，これらの関数の定数係数には差があるので，実際にどの程度の定数

のときに高い確率で単一解を持つ式を出力するかを実験的に調べたので，それについても報告する．

以上のように G を用いた場合，$m = \Theta(n \ln n)$ とすることで高い確率で単一解を持つ式が生成されることがわかったが，$m = \Theta(n \ln n)$ という項数はかなり大きいものであると考えられる．実際には，$m \approx 4.3n$ の項数を境に，3CNF 論理式の全体に対する充足可能な式の割合は急激に減少するので，$m \approx 4.3n$ 程度の論理式の例題が望まれている．そこで本論文では，より少ない項数で高い確率で単一解を持つ式を出力する手法として，G に対する次の 2 つの改良法を考えた．一つ目の方法は，2 つの充足解を保証した式をランダムに生成し，その充足解の片方を否定して単一解を持つ式にする方法である．充足解を 2 個持つ式は充足解を 1 個しか持たない式より少ない項数で構成できると考えられる．そこで，その 2 個の充足解の片方を否定する項を 1 つ追加すれば，充足解を 1 つに減らすことができると考えられる．もう一つの方法は，充足解の近傍の割当を否定する項を人為的に追加する方法である．ある充足解 t_0 の存在を保証してランダムに項を追加していくとき，最後まで充足解として残りやすい割当は t_0 の近傍に多いと思われるからである．これらの手法に対しても，高い確率で単一解を持つ式が生成されるための条件を理論的に解析し，また，実験的に定数の詳しい値を求めた．

2 準備

ここでは，基本的な用語や記号の定義を行う．

$(x_2 \vee \overline{x_4} \vee x_1) \wedge (\overline{x_3} \vee x_1 \vee x_2) \wedge (x_1 \vee x_2 \vee \overline{x_4})$ のような形の論理式を<u>3 和積標準形</u> (3-Conjunctive Normal Form, 3CNF) 論理式という．ここで，$x_1, \ldots$ を<u>変数</u>，x_i またはその否定 $\overline{x_i}$ を<u>リテラル</u>，$(x_2 \vee \overline{x_4} \vee x_1)$ のように 3 つのリテラルの論理和からなる式を<u>項</u>という．ただし，本論文では以下のような条件を満たすものとする．

- 項はちょうど 3 個のリテラルで構成される．
- 各項の中に同じ変数は高々 1 度しか出現しない．(例えば $(x_1 \vee x_1 \vee x_2)$ や $(x_1 \vee x_2 \vee \overline{x_1})$ のようなものは項としては扱わない．)
- 各項の中のリテラルの順序は問わない．(例えば $(x_1 \vee x_2 \vee x_3)$ と $(x_2 \vee x_3 \vee x_1)$ は同じ項として扱う．)
- 項の順序が異なる式は違う式とする．(例えば $(x_1 \vee x_2 \vee x_3)(x_4 \vee x_5 \vee x_6)$ と $(x_4 \vee x_5 \vee x_6)(x_1 \vee x_2 \vee x_3)$ は違う式である．)
- 項の重複を許す．(例えば $(x_1 \vee x_2 \vee x_3)(x_1 \vee x_2 \vee x_3)$ のような式も扱う．)

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(真理値)<u>割当</u>という．i 番目の変数 x_i の t における値は t の i ビット目(t_i)で表される．また，割当 t に対する f の値 $f(t)$ は f の各変数 x_i に t における値を代入して得られる値である．また，f を充足する(値が $f(t) = 1$ となる)割当を f の<u>充足割当</u>という．

任意の n 変数の真理値割当 t に充足される項の集合を C_t とする．3 個の変数の肯定・否定の組み合わせのうち 1 通りだけが t で充足できないので，$\|C_t\| = (2^3 - 1)\binom{n}{3} = 7\binom{n}{3}$ である．(以下では $\|C_t\| = N$ とする．)

論理式にある項 c を追加したときに，c によって初めてある割当 t' が充足解でなくなった場合 c が t' を<u>ヒット</u>したという．

適当な変数変換を施すことにより，G によって充足解 $\hat{t}$ に選ばれる割当は一般性を失うことなく 1^n を選んだものと仮定してよい．以下では，この割当を t_0

と表すこととする t_0 からハミング距離 k の割当の集合を T_k と表す また 与えられた n, m に対し このアルゴリズム G が出力する式を $f_{G(n,m)}$ と表記する これは， 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．

3 アルゴリズム G の解析

まず，アルゴリズム G が単一解を持つ式を高い確率で出力するときの 項数の上界に関する解析を行う．

定理 3.1 与えられた任意の $0 < \delta < 1$ に対して $m \geq \frac{14}{3} n \ln n - \frac{7}{3} n \ln \delta + o(1)$ のとき $G(n, m)$ は $1 - \delta$ の確率で単一解を持つ式を出力する．

証明 .まず，アルゴリズム G で生成された式 $f_{G(n,m)}$ が t_0 の他に t_0 から距離 k の t_k を解として持つ場合について考える このとき，$f_{G(n,m)}$ は $C_{t_0} \cap C_{t_k}$ の要素だけから構成されている $N_k = \|C_{t_0} \cap C_{t_k}\|$ とすると，$N_k = 6\binom{n}{3} + \binom{n-k}{3}$ であるので，$f_{G(n,m)}$ が t_0 の他に t_k を解として持つ確率は $\left\{\frac{N_k}{N}\right\}^m$ となる．したがって，$f_{G(n,m)}$ が t_0 以外に解を持つ確率を考えると，

$$\begin{aligned}
&\Pr_G\left[f_{G(n,m)} \text{ が } t_0 \text{ 以外に解を持つ}\right] \\
= &\Pr_G\left[\bigvee_{t \neq t_0} \left\{f_{G(n,m)} \text{ が } t \text{ も解に持つ}\right\}\right] \\
\leq &\sum_{t \neq t_0} \Pr_G\left[f_{G(n,m)} \text{ が } t \text{ も解に持つ}\right] \\
= &\sum_{k=1}^{n} \sum_{t_k \in T_k} \Pr_G\left[f_{G(n,m)} \text{ が } t_k \text{ も解に持つ}\right] \\
= &\sum_{k=1}^{n} \sum_{t_k \in T_k} \left\{\frac{N_k}{N}\right\}^m = \sum_{k=1}^{n} \binom{n}{k} \left\{\frac{N_k}{N}\right\}^m
\end{aligned}$$

となる この値が δ より小さければ

$$\Pr_G\left[f_{G(n,m)} \text{ が } t_0 \text{ だけを解に持つ}\right] \geq 1 - \delta$$

となる よってこの時の m の条件を求めることが目標となる 以下では，$p(k) = \binom{n}{k}\left\{\frac{N_k}{N}\right\}^m$ とおく．

まず，$p(k)$ が $1 \leq k \leq n$ で単調減少となる条件について考える $k > n - 3$ の場合は $N_{n-2} = N_{n-1} = N_n = 6\binom{n}{3}$ なので m によらず $p(k)$ は単調減少となる そこで $k \leq n - 3$ の場合について $\frac{p(k)}{p(k+1)} \geq 1$ となるための m の条件を考えると，

$$m \geq \frac{\ln \frac{n-k}{k+1}}{\ln\left(1 + \frac{\binom{n-k-1}{2}}{6\binom{n}{3} + \binom{n-k-1}{3}}\right)} \tag{1}$$

が得られる ここで，$\frac{n-k}{k+1}$ は $0 \leq k \leq n$ で単調減少，$1 + \frac{\binom{n-k-1}{2}}{6\binom{n}{3} + \binom{n-k-1}{3}}$ は $0 \leq k \leq n - 3$ で単調増加となるので 式 (1) の右辺は $0 \leq k \leq n$ で単調減少となり，$k = 0$ で最大値 $\frac{\ln n}{\ln \frac{n}{n - \frac{3}{7}}}$ をとる したがって，

$$m \leq \frac{\ln n}{\ln \frac{n}{n - \frac{3}{7}}} \tag{2}$$

のとき，$p(k)$ は $0 \leq k \leq n$ で単調減少となる つぎに，$p(1) < 1$ となる m の条件を考える． $p(1) = n\{\frac{6\binom{n}{3} + \binom{n-1}{3}}{7\binom{n}{3}}\}^m < 1$ より，

$$m > \frac{\ln n}{\ln \frac{n}{n - \frac{3}{7}}} \tag{3}$$

となる．

式 (2),(3) より $m > \frac{\ln n}{\ln \frac{n}{n - \frac{3}{7}}}$ のときに $p(k)$ は $0 \leq k \leq n$ で単調減少でかつ 1 以下の値を取る したがって $p(1) \leq \frac{\delta}{n}$ であれば，$\sum_{k=1}^{n} p(k) \leq \delta$ となる このためには，$m \geq \frac{2 \ln n - \ln \delta}{\ln \frac{n}{n - \frac{3}{7}}}$ が必要となるが 補題 3.2 より

$$m \geq \frac{14}{3} n \ln n - \frac{7}{3} n \ln \delta + o(1)$$

となる．□

補題 3.2 $\ln \frac{n}{n-d} = \frac{d}{n} + o(1)$.

証明 .$n \ln \frac{n}{n-d}$ を考える $\frac{1}{2} \ln \frac{1+x}{1-x} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$ であるので， $x = \frac{d}{2n-d}$ とすると，$n \ln \frac{n}{n-d} \to d$ となる よって，$\ln \frac{n}{n-d} = \frac{d}{n} + o(1)$. □

次に，アルゴリズム G が高い確率で単一解を持つ式を出力するときの 項数の下界に関する解析を行う．

定理 3.3 与えられた任意の $\varepsilon > 0$ に対して $m < \frac{7}{3}n\ln n + (\frac{7}{3}c - \alpha)n$ のとき $G(n,m)$ が複数解を持つ式を出力する確率は $1-\varepsilon$ 以上になる(ただし，c はオイラーの定数)よって十分大きな n に対して $H_n = \ln n + c$ とできるので

$$E[X] \approx \frac{7n}{3}(\ln n + c)$$

となる．一方 この幾何分布の分散を考えると

$$\sigma_{X_i}^2 = \frac{2-p_i}{{p_i}^2} = \frac{2\cdot 49n^2}{9(n-i)^2} - \frac{7n}{3(n-i)}$$

なので，

$$\begin{aligned}
\sigma_X^2 &= \sum_{i=0}^{n-1}\sigma_{X_i}^2 = \frac{2\cdot 49n^2}{9}\sum_{i=0}^{n-1}\frac{1}{(n-i)^2} - \frac{7n}{3}\sum_{i=0}^{n-1}\frac{1}{n-i} \\
&= \frac{2\cdot 49n^2}{9}\sum_{j=1}^{n}\frac{1}{j^2} - \frac{7n}{3}\sum_{j=1}^{n}\frac{1}{j} \\
&\leq \frac{2\cdot 49n^2}{9}\frac{\pi^2}{6} - \frac{7n}{3}\sum_{j=1}^{n}\frac{1}{j} < \frac{49\pi^2}{27}n^2
\end{aligned}$$

となる．すなわち $\sigma_X < \frac{7\pi}{3\sqrt{3}}n$ である また，

$$\begin{aligned}
&\Pr_G[m \text{ 項追加したとき解が } t_0 \text{ のみになる}] \\
\leq\ &\Pr_G[\forall t_1^i \in T_1 \text{ がヒットされる}] \\
=\ &\Pr_G[X \leq m] \\
=\ &\Pr_G[X - E[X] \leq m - E[X]] \\
\leq\ &\Pr_G[|X - E[X]| \geq E[X] - m]
\end{aligned}$$

となる．そこで，$E[X] - m \geq \sqrt{\frac{1}{\varepsilon}}\sigma_X$ となるように m を選び， Chebyshev の不等式 $\Pr_G[|X - E[X]| \geq t\sigma_X] \leq \frac{1}{t^2}$ を利用すると，

$$\begin{aligned}
&\Pr_G[|X - E[X]| \geq E[X] - m] \\
\leq\ &\Pr_G\left[|X - E[X]| \geq \sqrt{\frac{1}{\varepsilon}}\sigma_X\right] \leq \varepsilon
\end{aligned}$$

証明．この証明では T_1 の要素，すなわち 1bit だけ 0 を含むような割当を考える t_1^i を x_i だけが 0 である割当とする t_1^i をヒットする項は $(x_i \vee \overline{x_j} \vee \overline{x_k})$ となる項である(ただし $i \neq j \neq k$)このため， 1 つの項によって同時に異なる t_1^i と t_1^j がヒットされることはなく，t_1^i をヒットする項は C_{t_0} には $\binom{n-1}{2}$ 種類あることになる．

X をすべての T_1 の要素がヒットされるまでに追加する項の数に関する確率変数とし，$c_1, c_2, \ldots$ をランダムに追加する項の列とする(つまり c_X は最後に残った T_1 の要素をヒットする．)この列を，c_1 から最初に T_1 の要素をヒットした項 c_{i_1} までを区間 0，c_{i_1+1} から 2 回目に T_1 の要素をヒットした項 c_{i_2} までを区間 1 以下区間 $n-1$ までに分割する 区間 i に含まれる項の数に関する確率変数を X_i とする．($\sum_{i=0}^{i=n-1} X_i = X$ である．)

また，Y_i を t_1^i がランダムな項によってヒットされる事象とする．すると常に $\Pr_G[Y_i] = \frac{\binom{n-1}{2}}{7\binom{n}{3}} = \frac{3}{7n}$ となる また，一つの項が二つ以上の異なる t_1^i をヒットすることはないので，$i \neq j$ に対して Y_i と Y_j は排反事象である したがって区間 i において ランダムな項を追加したとき，その項が区間 i までにヒットされていない $n-i$ 個の T_1 の要素のうちどれか一つをヒットする確率は $\frac{3(n-i)}{7n}$ となる．つまり X_i は成功確率 $p_i = \frac{3(n-i)}{7n}$ となる幾何分布となる この幾何分布

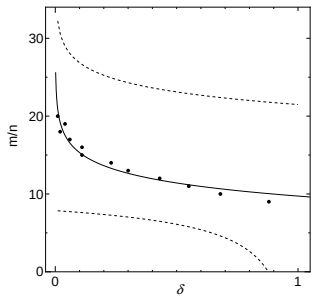

図 1: $n = 100$ のとき $G(n, m)$ が複数解を持つ式を出力する割合 δ と m/n の関係

となる このとき $E[X] - m = \alpha n \geq \sqrt{\frac{1}{\varepsilon}}\sigma_x$ の成立が必要となるが $\sigma_x < \frac{7\pi}{3\sqrt{3}}n$ であったので $\alpha n \geq \frac{7\pi}{3\sqrt{3\varepsilon}}n$ が成立すればよい したがって，

$$\alpha \geq \frac{7\pi}{3\sqrt{3\varepsilon}}$$

となる．□

これらの定理により示された項数の上界と下界には差が存在するので どの程度の項数が必要なのかを実験的に調べた 具体的には，$n = 100$ の場合について，m/n が 10 から 20 までの 1 毎に $G(n, m)$ が単一解を持つ式を出力した割合 $1 - \delta$ を調べた なお 単一解を持つかどうかの判定には ヒューリスティックな方法を用いた この結果を図 1 に示す .この結果を最小二乗法を用いて $a \ln 100 - b \ln \delta$ に近似をすると，$\frac{m}{n} = 9.73 - 2.43 \ln \delta = 2.11 \ln 100 - 2.43 \ln \delta$ (実線)が得られた なお 理論的解析で得られた値を破線で示してある．

4 必要な項数を減らす試み

3 節で，アルゴリズム G は 100 変数のときには 項変数比 m/n が 15 以上のときに高い割合で単一解を持つ式を出力することを示した しかし この項の数では項の数が多すぎるので 簡単に解を見つけられてしまうことが予想される(直感的には クロスオーバーポイントである $m \approx 4.3n$ に近いほうが難しいと思われる .)実際 この項の数で生成した例題はローカルサーチを利用した充足解探索アルゴリズムによって，ほとんどの場合 簡単に解である 1^n が発見されてしまう そこで 本節ではアルゴリズム G に改良を加え，より少ない項の数で単一解を持つ例題を生成することを考える．

まず，1 つめの方法として，2 つの解を持つ式を生成し その片方の解を否定する項を 1 つ追加する方法を考えた 具体的には 次のようなアルゴリズムである．

アルゴリズム G1

input: 変数の数 n, 項の数 m, 2 解の距離 d;
begin
　割当 $\hat{t}_1 \in \{0, 1\}^n$ を等確率で生成;
　割当 $\hat{t}_2$ を $d(\hat{t}_1, \hat{t}_2) = d$ となる割当から
　等確率で生成;
　for $i = 1$ **to** $m - 1$
　　do $\hat{t}_1, \hat{t}_2$ で充足される項を等確率で
　　生成し，f の i 番目の項として追加;
　$\hat{t}_1$ で充足され，$\hat{t}_2$ で否定される項を等確率で
　生成し，f の m 番目の項として追加;
　$(f, \hat{t}_1)$ を出力;
end.

アルゴリズム G1 では，2 解の距離 d によって必要な項数が変化することが予想される そこで まず $d = n$ のときに対して理論的な解析を行い 次に d を変化させたときの振舞いを実験的に調べた．

まず，G1 が高い確率で単一解を持つ式を出力する項数の上界について 定理 3.1 と同様にして次の定理が証明できる．

定理 4.1 与えられた任意の $0 < \delta < 1$ に対して $m \geq 4n \ln n - 2n \ln \delta + o(1)$ のとき $G1(n, m, n)$ は

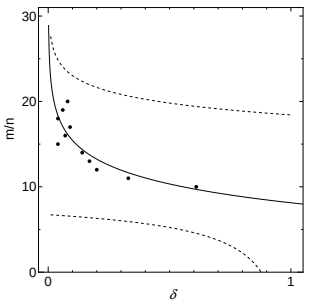

図 2: $n = 100$, $d = 100$ のとき $G1(n, m, d)$ が複数解を持つ式を出力する割合 δ と m/n の関係

$1 - \delta$ の確率で単一解を持つ式を出力する．

また 項数の下界に関しても 定理 3.2 と同様にして次の定理が証明できる．

定理 **4. 2** 与えられた任意の $\varepsilon > 0$ に対し，$m < 2 \ln n + (2c - \alpha)n$ のとき $G1(n, m, n)$ が複数解を持つ式を出力する確率は $1 - \varepsilon$ 以上になる(ただし，c はオイラーの定数であり，$\alpha \leq \frac{2\pi}{\sqrt{3\varepsilon}}$ である)．

これらの定理により示された項数の上界と下界には差が存在するので どの程度の項数が必要なのかを実験的に調べた 具体的には，$n = 100$ の場合について，m/n が 10 から 20 までの 1 毎に $G1(n, m, n)$ が単一解を持つ式を出力した割合 $1 - \delta$ を調べた なお 単一解を持つかどうかの判定には ヒューリスティックな方法を用いた この結果を図 2 に示す .この結果を最小二乗法を用いて，$a \ln 100 - b \ln \delta$ に近似をすると，$m/n = 8.1305012 - 3.18 \ln \delta = 1.77 \ln 100 - 3.18 \ln \delta$ (実線)が得られた なお 理論的解析で得られた値を破線で示してある この結果より，G1 は G より少ない項数でも高い確率で単一解を持つ式を出力することが分かる．

また，d を変化させたときに 必要な項数がどう変化するかについて実験的に調べた 図 3 に n を 100 に固定し，$d = 25, 50, 75, 100$ とした場合の結果を示す .このグラフから明らかなように，d が n に近づくほど少ない項数でも 単一解を持つ式を高い割合で出力している．

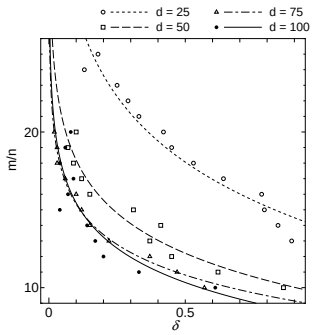

図 3: $n = 100$ で d を変えたときの m/n と δ の関係

次に，2 つめの改良方法として 解の近傍を否定する方法を考えた 具体的には次のようなアルゴリズムである．

アルゴリズム G2

input: 変数の数 n, 項の数 m;
begin
　真偽値割当 $\hat{t} \in \{0, 1\}^n$ を等確率で生成;
　for $i = 1$ **to** n
　　do $\hat{t}$ と i ビット目だけが異なる割当 t_1^i を否定する項を等確率で生成し，f の i 番目の項として追加;
　for $i = n + 1$ **to** m
　　do $\hat{t}$ で充足される項を等確率で生成し，f の i 番目の項として追加;
　$(f, \hat{t})$ を出力;
end.

アルゴリズム G2 に対しても 高い確率で単一解を持つ式を出力する項数の上界について解析を行い 定理 3.1 と同様にして次の定理を証明した．

定理 **4. 3** 与えられた任意の $0 < \delta < 1$ に対して，$m \geq \frac{7}{3} n \ln n + \left(\frac{7}{6} \ln \frac{2}{\delta} + 1\right) n + o(1)$ のとき $G2(n, m)$

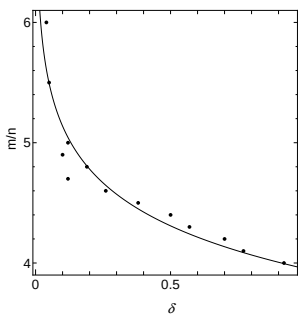

図 4: $n = 100$ のとき $G2(n, m)$ が複数解を持つ式を出力する割合 δ と m/n の関係

は $1 - \delta$ 以上の確率で単一解を持つ式を出力する．

また，実際にどの程度の項数が必要なのかを実験的に調べた．具体的には，$n = 100$ の場合について，$G(n, m)$ が単一解を持つ式を出力した割合 $1-\delta$ を調べた．なお，単一解を持つかどうかの判定には，ヒューリスティックな方法を用いた．この結果を図 4 に示す．この結果を最小二乗法を用いて，$a \ln 100 - b \ln \delta$ に近似をすると，$\frac{m}{n} = 3.95 - 0.515 \ln \delta = 0.858 \ln 100 - 0.515 \ln \delta$（実線）が得られた．なお，理論的解析で得られた値を破線で示してある．この結果から分かるように，100 変数の場合，ほぼ $m = 5n$ 程度の項数でも高い確率で単一解を出力していることが分かる．

5 おわりに

単純な 3SAT の例題生成アルゴリズムを解析し，高い確率で単一解を持つ式を出力する場合の項数の上界と下界を証明した．また，実験的に実験的に定数を調べた．

さらに，項数を減らしても高い確率で単一解を持つ式を出力するためにアルゴリズムに 2 種類の改良を行い，それぞれの場合について項数に関する解析を行い，オーダーとしては変らないものの係数が減少していることを示した．特に，2 つめの改良手法では 100 変数の場合，項の数が変数の数の 5 倍程度で高い確率で単一解を持つ式を出力することを実験的に示した．

本論文で扱った 3 つのアルゴリズムが高い確率で単一解を持つ式を出力したとき，最初の 2 つの生成手法で生成された式はローカルサーチを用いた充足解探索アルゴリズムで解を発見することが容易であるようである．しかし，最後の生成手法で生成された式はローカルサーチを用いた充足解探索アルゴリズムでは解を発見することが難しいようである．すなわちローカルサーチに対して難しい例題の生成手法となる可能性がある．こうした点の理論的・実験的解析が今後の重要な課題である．また $o(n \ln n)$ 個，特に $\Theta(n)$ 個の項を追加すれば高い確率で単一解になる問題を生成するようなアルゴリズムを作成することも重要な課題である．

参考文献

[1] 元木 光雄，充足不可能な論理式例題の生成に関する研究，卒業論文，東京工業大学，1996．

[2] 望月厚, 元木光雄, 渡辺治. 3 充足可能性判定問題 3SAT の正例題生成手法について. 電子情報通信学会技術研究報告, Vol. 96, No. 95, pp. 9–16, 1996.