ISSN 1814-1196
Научный вестник НГТУ
том 58, № 1, 2015, с. 183–200

http://journals.nstu.ru/vestnik
Science Bulletin of the NSTU
Vol. 58, No. 1, 2015, pp. 183–200

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

MODERN INFORMATION TECHNOLOGIES

УДК 004.023

Проектирование вещественных переменных в булевы в методе простой итерации, примененном к задаче выполнимости булевых формул*

Ю.Ю. ОГОРОДНИКОВ

644050, РФ, г. Омск, пр. Мира, 11, Омский государственный технический университет, аспирант. E-mail: yogorodnikov@gmail.com

Данная работа посвящена модификации непрерывного метода поиска решения задачи выполнимости булевых формул (SAT). Показан способ, по которому строится функционал, ассоциированный с SAT, и приведен общий алгоритм его решения методом простой итерации. Приведены ссылки на предыдущие работы автора и пояснено, что не всегда удается найти точное решение из-за попадания траектории метода простой итерации в овраг. Вместо построения эвристик по преодолению овражной ситуации и продолжения поиска предлагается сконструировать способ, который позволит определять некоторые биты приближения с высокой вероятностью. Полученные результаты могут быть с успехом использованы в задачах криптоанализа асимметричных шифров, логистики, автоматическом тестировании, распознавании данных, авторизации, в различных задачах на графах и других проблемах, которые сводятся к задаче SAT полиномиальным алгоритмом. Особое внимание уделяется задаче факторизации целых чисел, на которой построен известный асимметричный алгоритм шифрования RSA. Известно, что часть выполняющего набора SAT является ключом шифрования RSA. Соответственно, достаточно уметь определять часть выполняющего набора с высокой вероятностью. В предыдущих работах автора были определены некоторые нулевые биты с высокой вероятностью, в этой же работе основной упор делается на верное распознавание единичных бит. Для повышения числа единичных бит произведено построение нетривиального проектирования вещественных переменных в булевы. Представлено четыре способа: проектирование по окончании всех итераций, проектирование между итерациями с фиксированным параметром, проектирование с различными параметрами, байесовский подход. Проведено тестирование представленных методов на экземплярах 3-КНФ, эквивалентных задаче факторизации целых чисел. Полученные результаты сравниваются между собой, при этом критерием оптимальности выступает число верно определенных бит (как нулевых, так и единичных). Введены случайные величины ξ_0 и ξ_1, равные числу верно определенных нулевых и единичных бит соответственно. С помощью метода хи-квадрат Пирсона показано, что ξ_0 и ξ_1 имеют равномерное распределение, что позволяет выбирать произвольное стартовое приближение.

* *Статья получена 26 сентября 2014 г.*

Ключевые слова: метод простой итерации, выполнимость булевой формулы, проектирование вещественных переменных в булевы, параметр проектирования, байесовский подход, число верно определенных бит, равномерное распределение, единичные биты

DOI: 10.17212/1814-1196-2015-1-183-200

ВВЕДЕНИЕ

Задача выполнимости булевых формул (SAT или ВЫП) имеет широкий спектр приложений – от генерации автоматических тестов и составления расписания до задач криптографии [1]. Этим объясняется повышенное внимание к ней со стороны ученых-исследователей различных областей науки и техники [2–4]. К сожалению, в настоящее время неизвестно, существует или нет полиномиальный алгоритм решения SAT [5], однако для некоторых классов КНФ разработаны эффективные алгоритмы, способные за полиномиальное время дать ответ на поставленный вопрос [2, 6].

В некоторых случаях необязательно полностью искать выполняющий набор для формулы. Достаточно найти такой набор переменных, при котором число конъюнктов, принимающих значение ИСТИНА, будет максимальным. Данная задача получила название MAXSAT [7] и используется, например, в задаче формирования сложных изделий [8]. Другой задачей, не требующей полного нахождения выполняющего набора, является задача криптоанализа асимметричных шифров [9]. Стоит отметить, что данная задача допускает использование более широких методов оптимизации, в частности непрерывной оптимизации [10–12]. Р.Т. Файзуллиным разработан метод построения и минимизации непрерывного гладкого функционала на множестве вещественных чисел, соответствующего задаче SAT [13–15]. При тестировании уделялось особое внимание экземплярам SAT, эквивалентным задаче факторизации целых чисел. На последней задаче, как известно, основан алгоритм асимметричного шифрования RSA [16, 17].

К сожалению, разработанный метод не всегда находит точное решение, однако с его помощью возможно с высокой вероятностью определять некоторые биты выполняющего набора. Наибольший интерес среди них имеют так называемые биты ключа, т. е. позиции в выполняющем наборе, соответствующие ключу шифрования. В результате проведения статистического исследования [15] был получен массив вещественных переменных $(\nu_1, \nu_2, ..., \nu_n)$ (n – число переменных в SAT). Каждый элемент ν_i есть вероятность верного определения i-го бита. Для некоторых бит вероятность довольно высока ($\geq 0{,}95$), но большинство таких бит нулевые. Для успешного осуществления криптоанализа необходимо с высокой вероятностью уметь определять как единичные, так и нулевые биты [15].

В исследованиях [13–15] основное внимание уделяется изменению порядка вычислений переменных, выходу из «овражных» ситуаций, в которых оказывается траектория метода простой итерации, и статистическим обработкам полученных данных. Вопросу соотношения булевых и вещественных переменных уделено сравнительно немного внимания. В данной статье приводится описание эвристических методов преобразования вещественных переменных в булевы и обратно, сутью которых является повышение числа верно определяемых бит.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть на множестве переменных $y=(y_1,y_2...y_n)\in B^N$ задана форма $L^*(y)=\underset{i=1}{\overset{M}{\wedge}} G_i^*(y)$, где $G_i^*(y)=\underset{j=1}{\overset{N}{\vee}} q_{i,j}^*(y)$,

$$q_{i,j}^*(y)=\begin{cases} y_j, & \text{если переменная } y_j \text{ входит в } i\text{-й дизъюнкт непосредственно,} \\ \overline{y_j}, & \text{если переменная } y_j \text{ входит в } i\text{-й дизъюнкт с отрицанием,} \\ \textit{ложь, если переменная } y_j \textit{ не входит в } i\text{-й дизъюнкт.} \end{cases} \quad (1)$$

Преобразуем 3-КНФ в 3-ДНФ, т. е. каждый дизъюнкт заменим на конъюнкт по правилам:

$$L(y)=\overline{L^*(y)}=\underset{i=1}{\overset{M}{\vee}} G_i(y) \quad (2)$$

$$G_i(y)=\underset{j=1}{\overset{N}{\wedge}} q_{i,j}(y) \text{ и } q_{i,j}(y)=\overline{q_{i,j}^*(y)}.$$

Целесообразность такого перехода продиктована тем, что вычислительные операции, необходимые для работы с 3-ДНФ, занимают меньше процессорного времени и более удобны для программирования [14]. Также известно, что набор булевых переменных y_j, обращающий формулу (1) в значение ИСТИНА, обращает 3-ДНФ, полученную по правилам (2), в значение ЛОЖЬ [14].

Далее преобразуем булевы переменные y_j в вещественные x_j по следующим соотношениям:

$$x_j=\begin{cases} 1, \text{ если } y_j=\textit{истина}, \\ 0, \text{ если } y_j=\textit{ложь}. \end{cases} \quad (3)$$

Следующим шагом сопоставим форме $L(y)$ функционал $F:[0,1]^N\to R_+$ следующего вида:

$$F(x)=\sum_{i=1}^{M} C_i(x),$$

$$\text{где } C_i(x)=\prod_{j=1}^{N} p_{i,j}(x) \text{ и } p_{i,j}(x)=\begin{cases} x_j^2, & \text{если } q_{i,j}^*(y)=\overline{y_j}, \\ (1-x_j)^2, & \text{если } q_{i,j}^*(y)=y_j, \\ 1, & \text{если } q_{i,j}^*(y)\neq\overline{y_j} \text{ и } q_{i,j}^*(y)\neq y_j. \end{cases} \quad (4)$$

Таким образом, получен функционал, заданный многочленом степени $2k$, где k – максимальное число литералов, входящих в один конъюнкт

$p_{i,j}(x)$. Так, для случая 3-ДНФ степень функционала равняется шести, так как число литералов, входящих в один дизъюнкт, равняется трем [15]. Известно, что глобальный минимум $F(x)$ соответствует набору булевых переменных y_j, обращающих $L(y)$ в значение ЛОЖЬ, а $L^*(y)$ – в ИСТИНУ [14].

Будем искать глобальный минимум $F(x)$. Вместо традиционных методов будем искать стационарные точки функционала. Для этого дифференцируем $F(x)$ по всем переменным, тем самым получим его градиент ∇F, и приравниваем к нулю его компоненты [18]. Получаем систему уравнений вида

$$\frac{\partial F}{\partial x_i} = 0. \tag{5}$$

Распишем систему (5). Так как при дифференцировании функционала $F(x)$ по *i*-й переменой слагаемые $C_j(x)$, которые не содержат x_i, обращаются в ноль, то уравнение (5) можно записать в виде

$$x_i \sum Q_i(x) - (1 - x_i) \sum \overline{Q_i(x)} = 0. \tag{6}$$

Здесь $Q_i(x)$ – дифференцированное слагаемое, содержащее переменную x_i без отрицания, $\overline{Q_i(x)}$ – аналогичное слагаемое, только для случая, когда x_i входит с отрицанием. Например, пусть имеются два слагаемых, содержащих переменную x_1:

$$C_1 = x_1^2 x_2^2 x_3^2 \text{ и } C_2 = (1 - x_1)^2 x_2^2 (1 - x_3)^2.$$

Тогда $Q_1(x) = x_2^2 x_3^2$ и $Q_2(x) = x_2^2 (1 - x_3)^2$.

Раскроем скобки в (5) и выразим переменную x_i:

$$x_i = \frac{\sum \overline{Q_i(x)}}{\sum Q_i(x) + \sum \overline{Q_i(x)}}, \quad i = 1...N. \tag{7}$$

Будем решать получившуюся систему методом простой итерации. В качестве стартового приближения в общем случае можно выбрать вектор $x^{(0)}$, равномерно заполненный значениями 0 и 1, однако эксперименты показали, что метод лучше сходится к решению, если взять стартовое приближение с произвольными значениями из множества натуральных чисел.

Также следует сделать важное замечание касательно вещественных переменных x_i. Как нетрудно заметить, из соотношения (7) следуют ограничения $0 \le x_i \le 1$. Действительно, рассмотрим слагаемые $\sum Q_i(x)$ и $\sum \overline{Q_i(x)}$. Каждое из них является неотрицательным числом. Случаи, когда оба числа равняются нулю, обрабатываются в порядке исключения, тогда переменной x_i присваивается значение 0,5. Однако в экземплярах SAT, использованных при тестировании методов данной статьи, данного случая не возникает в силу особенностей исходной 3-КНФ.

Пусть далее $\sum\overline{Q_i(x)}$ не равно нулю. Тогда можно сделать преобразование

$$x_i = \frac{\sum\overline{Q_i(x)}}{\sum Q_i(x) + \sum\overline{Q_i(x)}} = \frac{1}{\frac{\sum Q_i(x)}{\sum\overline{Q_i(x)}} + 1}. \quad (8)$$

Отсюда следует, что $x_i \le 1$. Причем равенство достигается в случае, когда $\sum Q_i(x) = 0$, т. е. когда переменная y_i, соответствующая x_i, входит в $L^*(y)$ только с отрицанием. В остальных же случаях $x_i < 1$.

Ниже приведен псевдокод метода.

Алгоритм 1. Метод простой итерации минимизации функционала, ассоциированного с SAT.

Входные данные: вектор-приближение $x^{(0)}$.

Шаг 1. Сделать $x^{(0)}$ текущим приближением.

Шаг 2. Положить $i = 1$ (счетчик числа итераций).

Шаг 3. Положить $j = 1$ (счетчик по переменным).

Шаг 4. Вычислить значение x_i по формуле (8).

Шаг 5. Добавить компоненту x_i к новому приближению $x^{(i)}$.

Шаг 6. Если $j < N$, где $N -$ это число переменных, то перейти к шагу 3.

Шаг 7. Сделать приближение $x^{(i)}$ текущим.

Шаг 8. Если $i < \text{count\_iter}$, где count_iter – это количество итераций в методе, то перейти к шагу 2.

Выходные данные: приближенное решение $x^{(\text{count\_iter})}$.

В качестве выходных данных выступает вещественный вектор $x^{(\text{count\_iter})}$. Естественно, он не является выполняющим набором для формы $L^*(y)$, и требуется провести проектирование вещественных переменных в булевы. Другими словами, требуется выполнить операцию, обратную действиям (3).

2. ПРОЕКТИРОВАНИЕ ВЕЩЕСТВЕННЫХ ПЕРЕМЕННЫХ В БУЛЕВЫ

В данной главе представлены описания и результаты исследования различных способов проектирования. Тестирование проводилось на экземплярах SAT, эквивалентной задаче факторизации целого числа размерности 100. Тестируемая 3-КНФ содержит $N = 14\ 700$ переменных и $M = 58\ 200$ дизъюнктов.

3. ПРОЕКТИРОВАНИЕ ПОСЛЕ ОКОНЧАНИЯ ИТЕРАЦИЙ

Рассмотрим простейший вариант – параметр проектирования (будем обозначать его буквой θ) фиксирован и равен вещественному числу из диа-

пазона $(0...1)$. При этом проектирование проводится только по завершении итераций по соотношению

$$y_j = \begin{cases} \text{истина} & x_j \le \theta, \\ \text{ложь} & x_j > \theta. \end{cases} \tag{9}$$

В этом случае вне зависимости от θ метод будет приходить в одну и ту же точку пространства поиска R^{N}. Остается лишь выбрать значение параметра θ, соответствующее максимальному числу верно определяемых бит. В табл. 1 приведены данные по исследованию данного случая.

Таблица 1

Проектирование по окончании выполнения всех итераций

Значение параметра θ	Число верно определенных единиц	Число верно определенных нулей	Общее число верно определенных бит
0.1	587	9559	10 146
0.2	561	9657	10 218
0.3	508	9857	10 365
0.4	477	10 005	10 482
0.5	458	10 034	10 492
0.6	454	10 057	10 511
0.7	426	10 110	10 536
0.8	399	10 229	10 628
0.9	369	10 292	10 661

4. ПРОЕКТИРОВАНИЕ ПОСЛЕ ВЫПОЛНЕНИЯ КАЖДОЙ ИТЕРАЦИИ

Рассмотрим модификацию предыдущего метода. Параметр θ также остается неизменным на протяжении выполнения алгоритма, но проектирование переменных будет проводиться не по окончании всех итераций, а после выполнения каждой из них. Другими словами, проектирование будет проводиться между шагами 6 и 8 алгоритма 1.

Поясним математический смысл данного метода. При старте метода простой итерации выбирается некоторое начальное приближение $x^{(0)}$. Оно вещественное и соответствует точке в пространстве поиска R^N. После выполнения одной итерации получается новая точка с вещественными координатами $x^{(1)}$. Так как для всех переменных x_i^k справедливо соотношение $0 \le x_i^k \le 1$, то можно сказать, что поиск ведется в N-мерном единичном кубе. Тогда задача проектирования в булево пространство фактически является

задачей выбора вершины единичного куба, которая лучше всего подходит для дальнейшего поиска. В табл. 2 представлены результаты исследования проектирования между итерациями.

Таблица 2

Проектирование между итерациями

Значение параметра θ	Число верно определенных единиц	Число верно определенных нулей	Общее число верно определенных бит
0.1	49	3527	3576
0.2	765	3298	4063
0.3	780	3370	4150
0.4	803	3480	4283
0.5	946	4201	5147
0.6	918	4408	5326
0.7	500	3320	3820
0.8	70	3150	3220
0.9	26	3217	3243

5. ПРОЕКТИРОВАНИЕ С НЕСКОЛЬКИМИ ПАРАМЕТРАМИ θ

Рассмотренные выше способы проектирования обладают одним существенным недостатком – параметр θ одинаков для всех переменных. В общем случае этот вариант не является эффективным.

Логичным было бы предположить, что для каждой переменной должен быть свой параметр θ_i. Более того, переменные можно объединять в группы по некоторому признаку и для каждой группы выделять свой параметр проектирования. Проблема заключается в том, что очень часто переменные трудно разбить на группы. Это может быть разбиение и по частоте появления в КНФ, и по зависимости с другими переменными и т. д. Этот подход схож с выбором весовых множителей для дизъюнктов [3] и носит эвристический характер.

В табл. 3 приведены результаты исследований проектирования с двумя параметрами θ. В качестве тестового примера, как отмечалось выше, была взята 3-КНФ, ассоциированная с задачей факторизации целых чисел. При этом битам, отвечающим битам сомножителя, был сопоставлен параметр θ_1, остальным же – θ_2. В сумме эти параметры необязательно должны равняться единице. В представленных данных параметр θ_1 последовательно брался равным 0,1; 0,2; 0,3;…0,9, а θ_2 подбирался так, чтобы максимальное число верно определенных бит было максимальным при фиксированном θ_1.

Таблица 3

Результаты исследования проектирования с двумя параметрами θ_1 и θ_2

Значение параметра θ_1	Значение параметра θ_2	Число верно определенных единиц	Число верно определенных нулей	Общее число верно определенных бит
0.1	0.7	29	3537	3566
0.2	0.7	24	3644	3668
0.3	0.8	121	4012	4133
0.4	0.6	840	4105	4945
0.5	0.7	940	4153	5093
0.6	0.3	916	4148	5064
0.7	0.3	205	2365	2570
0.8	0.2	104	1827	1931
0.9	0.1	24	806	830

6. БАЙЕСОВСКИЙ ПОДХОД К ПРОЕКТИРОВАНИЮ ПЕРЕМЕННЫХ

Данный способ подразумевает под собой построение байесовского классификатора, который определяет наиболее вероятный класс распознавания [19]. Переформулируем задачу проектирования в терминах байесовского подхода. Обозначим через $x^{(t)}$ вектор-приближение, сформированный на t-й итерации. Тогда $x_i^{(t)}$ – t-й бит вектора-приближения. Введем два класса $\Omega_0 = \{x_i^t = 0\}$ и $\Omega_1 = \{x_i^t = 1\}$. Очевидно, что эти классы не пересекаются и образуют полное пространство классов $\Omega = \Omega_0 \cup \Omega_1$. Далее введем два события: $H_i^0 = \{x_i^t \in \Omega_0\}$ и $H_i^1 = \{x_i^t \in \Omega_1\}$. Эти события являются гипотезами и состоят в том, что t-й бит вектора-приближения принадлежит к классам Ω_0 и Ω_1.

Пусть A_i – событие, состоящее в том, что i-й бит вектора-приближения определен верно, т. е. совпадает с точным решением. Тогда $P(A_i)$ – вероятность этого события.

По формуле полной вероятности имеем

$$P(A_i) = P\left(H_i^0\right)P\left(A_i \mid H_i^0\right) + P\left(H_i^1\right)P\left(A_i \mid H_i^1\right).$$

Здесь $P\left(A_i \mid H_i^0\right)$ – вероятность верного определения *i*-го бита при условии, что он отнесен к классу Ω_0 (другими словами, распознан как нулевой бит). Соответственно, $P\left(A_i \mid H_i^1\right)$ – вероятность верного определения *i*-го бита при распознавании его как единичного.

По формуле Байеса имеем

$$P\left(H_i^0 \mid x_i^t\right) = \frac{P\left(H_i^0\right)P\left(x_i^t \mid H_i^0\right)}{P\left(x_i^t\right)} \text{ и } P\left(H_i^1 \mid x_i^t\right) = \frac{P\left(H_i^1\right)P\left(x_i^t \mid H_i^1\right)}{P\left(x_i^t\right)}, \quad (10)$$

$P\left(H_i^0 \mid x_i^t\right)$ и $P\left(H_i^1 \mid x_i^t\right)$ – это вероятности верного отнесения бита $x_i^{(t)}$ к классам Ω_0 и Ω_1 при условии, что $x_i^{(t)}$ верно определен. Именно на данных апостериорных вероятностях вычисляются параметры проектирования. Для успешного их вычисления требуется знание четырех вероятностей: $P\left(H_i^0\right)$, $P\left(H_i^1\right)$, $P\left(A_i \mid H_i^0\right)$, $P\left(A_i \mid H_i^1\right)$.

Вероятности $P\left(H_i^0\right)$ и $P\left(H_i^1\right)$ – безусловные. Изначально они задаются эвристически, например, можно взять $P\left(H_i^0\right) = 0.6$ и $P\left(H_i^1\right) = 0.4$. В процессе выполнения метода простой итерации вероятности $P\left(H_i^0\right)$ и $P\left(H_i^1\right)$ пересчитываются в зависимости от значения переменной x_i^t и параметра проектирования θ:

$$\begin{array}{ll} P(H_0) = x_i^t, \quad \text{если} \quad x_i^t \le \theta, & \\ P(H_0) = 1 - x_i^t, \quad \text{если} \quad x_i^t > \theta & \end{array} \quad \text{и} \quad \begin{array}{ll} P(H_1) = 1 - x_i^t, \quad \text{если} \quad x_i^t \le \theta, & \\ P(H_1) = x_i^t, \quad \text{если} \quad x_i^t > \theta. & \end{array}$$

Вероятности $P\left(A_i \mid H_i^0\right)$ и $P\left(A_i \mid H_i^1\right)$ – априорные. Они могут быть получены на основании статистических испытаний обыкновенного метода простой итерации. В работе [5] описаны методы по определению вероятности верного определения бит. К сожалению, лишь для некоторых нулевых бит удалось получить высокие вероятности, для остальных же они существенно ниже. Однако полученные данные можно использовать как априорные в байесовском проектировании.

Таким образом, после каждой итерации метода последовательных приближений происходит вычисление апостериорных вероятностей по формулам (10). В зависимости от того, какая вероятность окажется больше, происходит проектирование либо в ноль, либо в единицу. В табл. 4 приведены результаты тестирования данного подхода.

Таблица 4

Тестирование байесовского подхода к проектированию

Параметр проектирования θ	Число верно определенных единиц	Число верно определенных нулей	Общее число верно определенных бит
0,1	2324	1700	4024
0,2	2335	1704	4039
0,3	2314	1753	4067
0,4	3470	450	3920
0,5	3468	441	3909
0,6	3482	421	3903
0,7	2441	2170	4611
0,8	2292	2671	4963
0,9	2219	2641	4860

Этот подход, так же как и предыдущие, имеет свои недостатки. Самым главным из них является то, что он предполагает независимость всех переменных. В таком случае говорят о *наивном* байесовском классификаторе.

7. СРАВНЕНИЕ МЕТОДОВ ПРОЕКТИРОВАНИЯ

В данной главе приведено сравнение всех представленных методов проектирования. Критерием оптимальности выступает число верно определенных бит.

На рис. 1 изображен график зависимости числа верно определенных бит от θ. Для построения линии метода 2.3 учитывался только параметр θ_1, параметр θ_2 определяется для него однозначно из табл. 3.

Как нетрудно заметить, первый метод определяет максимальное число бит, однако большая часть из них нулевые. Это объясняется тем, что проектирование производится лишь после окончания выполнения всех итераций, после каждой из которых переменные x_i^t уменьшаются из-за погрешностей вычислений по формуле (8). Исключения составляют лишь те переменные, которые входят в функционал исключительно в отрицательном виде – они будут сходиться к единице. Остальные же будут сходиться к нулю. Становится очевидным, что данный метод не является оптимальным.

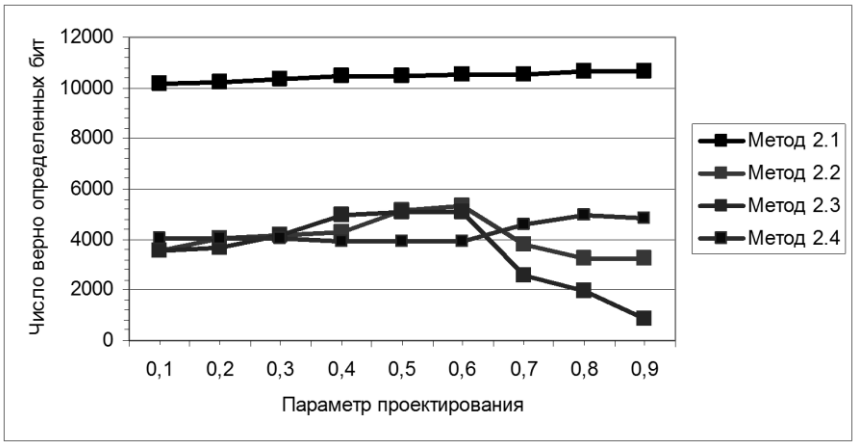

Рис. 1. График зависимости числа верно определяемых бит от параметра проектирования θ

Остальные методы определяют примерно одинаковое количество бит, байесовский подход даже меньше, чем остальные. Но у последнего есть существенное преимущество – он распознает больше единичных бит, чем остальные. На рис. 2 показан график зависимости числа верно определенных единичных бит от θ.

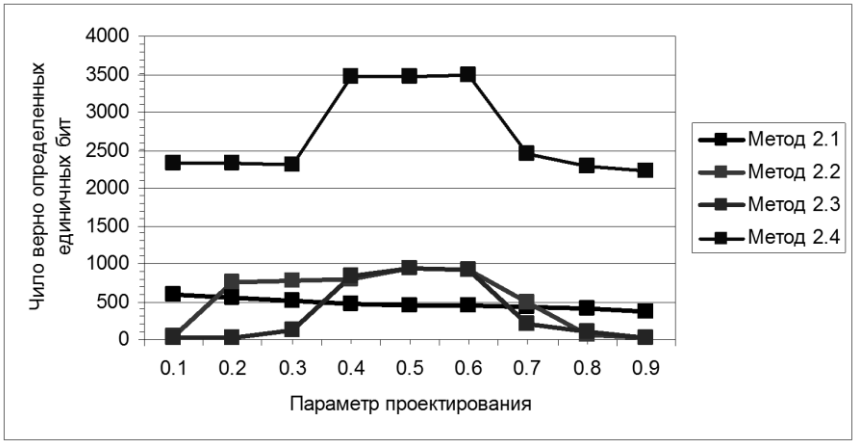

Рис. 2. График зависимости числа верно определенных единичных бит от параметра проектирования θ

Из графика видно, что оптимальный результат достигается при $0,4 \le \theta \le 0,6$. Причем число верно определенных единиц увеличивается за счет сокращения верно определенных нулей. Общее же число остается примерно одинаковым для любого $\theta \in \{0,1\}$. Данный результат является эвристическим, и на текущий момент строго не доказан. Интересным также представляется соотношение общего числа определенных единиц и числа верно определенных единиц, т. е. процент верно определенных бит. В табл. 5 приведены данные по процентному соотношению.

Таблица 5

Соотношение числа верно определенных единиц и общего числа определенных единиц

Номер метода	Максимальное число верно определенных единиц N_1	Общее число определенных единиц N_2	Отношение N_1 / N_2
2.1	587	3801	0.1544
2.2	946	5702	0.16509
2.3	940	5361	0.1753
2.4	3482	4890	0.712

Как нетрудно заметить из приведенной таблицы, общее число определенных бит при байесовском подходе не так велико по сравнению с остальными методами, однако большая часть из определенных единичных бит – верные. Таким образом, когда при выполнении проектирования по методу 2.4 определяются единичные биты, то мы можем утверждать, что они совпадают с верными с вероятностью, превышающей 71 %.

Покажем также, что число верно определенных бит не зависит от выбора начального приближения. Для этого исследуем полученные результаты с помощью методов математической статистики.

Обозначим через ξ_0 и ξ_1 случайные величины, равные числу верно определенных нулей и единиц. С помощью метода хи-квадрат Пирсона [20] проверим гипотезы о равномерном распределении ξ_0 и ξ_1. Для этого проведем эксперимент, состоящий в том, что на вход алгоритма 1 подаются различные стартовые приближения, и в результате подсчитывается число верно определенных нулей и единиц. В результате проведения эксперимента была получена выборка $(X_1, X_2, ...X_n)$ из $n = 500$ элементов.

Проверим гипотезу о равномерном распределении ξ_1. В получившейся выборке максимальный элемент оказался равным $x_{\max} = 4133$, минимальный – $x_{\min} = 4006$. Упорядочим значения тестируемой выборки по возрастанию и разобьем их на $l = 10$ интервалов равной длины $h = \dfrac{x_{\max} - x_{\min}}{1 + 3.222 \lg n} = 13$. Обозначим границы интервалов через a_i. Число точек, попавших в интервал $[a_i, a_{i+1}]$, обозначим через m_i. Теоретическая вероятность попадания случайной величины X в каждый интервал, согласно гипотезе о равномерном распределении, равна $p_i = \int\limits_i^{i+1} \frac{1}{l} dx = \frac{1}{l}(i + 1 - i) = \frac{1}{l} = \frac{1}{10}$. Тогда математическое ожидание числа попаданий в интервал $np_i = 500 \dfrac{1}{10} = 50$ для всех $i = 1...l$. Так как $np_i \geq 10$, то нет смысла в объединении интервалов разбиения. Дальнейшие вычисления приведены в табл. 6.

Таблица 6

Расчеты для проверки гипотезы о равномерном распределении ξ_1

Номер интервала	Границы интервала	m_i	np_i	$m_i - np_i$	$\frac{(m_i - np_i)^2}{np_i}$
1	[4006; 4019)	47	50	–3	0.18
2	[4019; 4032)	56	50	6	0.72
3	[4032; 4045)	48	50	–2	0.08
4	[4045; 4058)	45	50	–5	0.5
5	[4058; 4071)	46	50	–4	0.32
6	[4071; 4084)	45	50	–5	0.5
7	[4084; 4097)	48	50	–2	0.08
8	[4097; 4110)	57	50	7	0.98
9	[4110; 4123)	60	50	10	2
10	[4123; 4133]	48	50	–2	0.08
Σ		500	500		5.44

Таким образом, числовое значение $K_{\max} = 5.44$. Число степеней свободы примем равным $t = l - 1 = 9$. Для заданного уровня значимости $\alpha = 0.05$ находим $\gamma = 1 - \alpha = 0.95$, $\chi^2 = (0.95;\ 9) = 16.9190$. Так как $K_{\max} < \chi^2$, то гипотеза о равномерном распределении случайной величины ξ_1 принимается.

Покажем теперь, что случайная величина ξ_0 также имеет равномерное распределение (табл. 7). Пусть по-прежнему имеется выборка $(X_1, X_2, ...X_n)$ из $n = 500$ элементов. Проделаем те же шаги, что и для ξ_1. Получаем $x_{\min} = 394$ и $x_{\max} = 662$. Упорядочим элементы выборки по возрастанию и разобьем ее на $l = 10$ интервалов длины $h = \frac{x_{\max} - x_{\min}}{1 + 3.222 \lg n} = \frac{662 - 394}{1 + 3.222 \lg 500} = 27.$

Числовое значение $K_{\max} = 6$. Число степеней свободы, так же как и в предыдущем случае, примем равным $t = l - 1 = 9$. Для заданного уровня значимости $\alpha = 0.05$ находим $\gamma = 1 - \alpha = 0.95$, $\chi^2 = (0.95;\ 9) = 16.9190$. Так как $K_{\max} < \chi^2$, то гипотеза о равномерном распределении случайной величины ξ_0 принимается.

Таблица 7

Расчеты для проверки гипотезы о равномерном распределении ξ_0

Номер интервала	Границы интервала	m_i	np_i	$m_i - np_i$	$\frac{(m_i - np_i)^2}{np_i}$
1	[394; 421)	55	50	5	0.5
2	[421; 448)	53	50	3	0.18
3	[448; 475)	45	50	–5	0.5
4	[475; 502)	46	50	–4	0.32
5	[502; 529)	44	50	–6	0.72
6	[529; 556)	48	50	–2	0.08
7	[556; 583)	46	50	–4	0.32
8	[583; 610)	54	50	4	0.32
9	[610; 637)	62	50	12	2.88
10	[637; 662]	47	50	–3	0.18
Σ		500	500		6

С практической точки зрения равномерное распределение случайных величин ξ_0 и ξ_1 позволяет выбрать произвольное начальное приближение, состоящее из множества натуральных чисел.

ЗАКЛЮЧЕНИЕ

Задача выполнимости булевых формул имеет важное значение во многих областях науки и техники: в системе автоматической проверки тестов, задачах логистики, составлении расписаний, криптографии и криптоанализе. Особое внимание в статье уделяется задаче факторизации, так как на последней основан известный алгоритм шифрования RSA. Таким образом, если будет найден полиномиальный алгоритм решения задачи SAT, то дешифрование RSA также будет проводиться за полиномиальное время. К сожалению, в настоящее время неизвестен такой «быстрый» алгоритм, однако ученые всего мира продолжают различные алгоритмы для решения SAT. Одним из малоизученных подходов является построение вероятностного алгоритма, суть которого заключается в накоплении статистики верного определения бит и в последующем ее использовании. Преимуществом данного алгоритма является высокая скорость работы. Недостатком же является отсутствие гарантии,

что полученные данные всегда дают верный ответ. Однако данный недостаток можно устранить, увеличивая число бит, для которых частота совпадения с точными довольно высока. Другим вариантом использования статистики является ее использование в точных алгоритмах нахождения решения SAT.

В статье описан метод простой итерации, с помощью которого осуществляется накопление статистики верного определения бит. В предыдущих работах автора [13, 15] приведены данные по накоплению статистики верного определения бит. К сожалению, с высокой вероятностью (≥ 0.95) удалось определить только нулевые биты. В данной статье приводится модификация, призванная повысить число верно определяемых единичных бит. Модификация заключается в построении нетривиального метода проектирования вещественных переменных в булевы. Как видно из полученных результатов, наиболее оптимальным является байесовский подход – в результате его применения удается верно определить единичные биты с вероятностью 71 %. Также его особенностью является сокращение вероятности верного определения нулевых бит. Кроме того, было проведено исследование с помощью методов математической статистики и установлено, что число верно определяемых единичных и нулевых бит в случае байесовского проектирования является равномерной величиной. Этот факт позволяет брать произвольное стартовое приближение для метода простой итерации.

Полученные результаты могут быть использованы прежде всего в задаче криптоанализа RSA, но также возможно применение разработанного метода к широкому спектру задач, полиноимально сводящихся к SAT.

СПИСОК ЛИТЕРАТУРЫ

1. *Левин Л.А.* Универсальные задачи перебора // Проблемы передачи информации. – 1973. – Т. 9, вып. 3. – С. 115–116.

2. Sat Live! [Electronic resource]: website. – URL: www.satlive.org (accessed: 28.01.2015).

3. *Gu J.* Local search for satisfiability (SAT) problem // IEEE Transactions on Systems, Man, and Cybernetics. – 1993. – Vol. 23, iss. 4. – P. 1108–1129. – doi: 10.1109/21.247892.

4. *Gu J.* On optimizing a search problem // Artificial Intelligence Methods and Applications / N.G. Bourbakis, ed. – New Jersey: World Scientific Publishing, 1992. – Ch. 2. – P. 63–105. – (Advanced series on Artificial Intelligence; vol. 1).

5. *Cook S.A.* The complexity of theorem proving procedures // Proceedings of the Third Annual Symposium on Theory of Computing, STOC '71. – New York, USA: ACM Press, 1971. – P. 151–158. – doi: 10.1145/800157.805047.

6. *Davis M., Logeman G., Loveland D.* A machine program for theorem proving // Communication of the ACM. – 1962. – Vol. 5, N 7. – P. 394–397. – doi: 10.1145/368273.368557.

7. *Batiti R., Protasi M.* Approximate algorithms and heuristics for MAX-SAT // Handbook of Combinatorial Optimization. – Kluwer Academic Publishers, 1998. – Vol. 1. – P. 77–148. – doi: 10.1007/978-1-4613-0303-9_2.

8. *Гуселетова О.Н.* Математические модели и алгоритмы дискретной оптимизации для решения задач формирования сложных изделий: автореф. дис. … канд. техн. наук. – Омск, 2008. – 16 с.

9. Handbook of applied cryptography / A.J. Menezes, P.C. van Oorschot, S.A. Vanstone. – Boca Raton, Florida, USA: CRC Press, 2001. – 780 p.

10. *Маслов С.Ю.* Итеративные методы в переборной модели, как модель интуитивных // Тезисы IX Всесоюзного симпозиума по кибернетике, 10–15 ноября 1981 г. – Сухуми, 1981. – С. 26–28.

11. *Крейнович В.Я.* Семантика итеративного метода С.Ю. Маслова // Вопросы кибернетики: сборник статей. – М.: [б. и.], 1987. – Вып. 131: Проблемы сокращения перебора. – С. 30–62.

12. *Опарин Г.А., Новопашин А.П.* Непрерывные модели решения систем булевых уравнений // Вестник Томского государственного университета. Приложение: Материалы международных всесоюзных и региональных научных конференций, симпозиумов, школ, проходивших в ТГУ. – 2004. – № 9 (1). – С. 20–25.

13. *Файзуллин Р.Т., Дулькейт В.И., Огородников Ю.Ю.* Гибридный метод поиска приближенного решения задачи 3-выполнимость, ассоциированной с задачей факторизации // Труды Института математики и механики УрО РАН. – 2013. – Т. 19, № 2. – С. 285–294.

14. *Дулькейт В.И., Файзуллин Р.Т., Хныкин И.Г.* Непрерывные аппроксимации решения задачи «выполнимость» применительно к криптографическому анализу асимметричных шифров // Компьютерная оптика. – 2009. – Т. 33, № 1. – С. 68–73.

15. *Огородников Ю.Ю., Файзуллин Р.Т.* Определение нулевых бит задачи 3-выполнимость, ассоциированной с задачей факторизации // Компьютерная оптика. – 2014. – Т. 38, № 3. – С. 521–528.

16. *Rivest R.L., Shamir A., Adleman L.* A method for obtaining digital signatures and public-key cryptosystems // Communications of the ACM. – 1978. – Vol. 21, iss. 2. – P. 120–126. – doi: 10.1145/359340.359342.

17. *Patsakis C.* RSA private key reconstruction from random bits using SAT solvers [Electronic resource]. – Received 18 Jan 2013, last revised 7 May 2013. – (Cryptology ePrint Archive: report 2013/026). – URL: https://eprint.iacr.org/2013/026 (accessed: 28.01.2015).

18. *Березин И.С., Жидков Н.П.* Методы вычислений: учебное пособие для вузов. Т. 2. – М.: Физматлит, 1959. – Гл. 7, § 5. Решение систем уравнений. – С. 150–162.

19. *Местецкий Л.М.* Математические методы распознавания образов: курс лекций [Электронный ресурс] / МГУ, [факультет] ВМиК, кафедра «Математические методы прогнозирования». – М., 2002. – Гл. 2, § 2.1. Классификация на основе Байесовской теории решений. – С. 8–13. – URL: http://www.ccas.ru/frc/papers/mestetskii04course.pdf (дата обращения: 28.01.2015).

20. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами: учебное пособие / А.И. Кибзун, А.Р. Горяинова, А.В. Наумов, А.Н. Сиротин. – М.: Физматлит, 2002. – 224 с.

Огородников Юрий Юрьевич, аспирант кафдеры комплексной защиты информации радиотехнического факультета Омского государственного технического университета. Основное направление научных исследований – поиск полного и частичного решения задачи выполнимости булевых формул, распознавание образов, программирование математических расчетов. Имеет 7 научных публикаций. E-mail: yogorodnikov@gmail.com.

Projection of real variables to Boolean variables by the simple iteration method applied to the problem of the Boolean formula satisfiability[*]

Yu.Yu. OGORODNIKOV

Omsk State Technical University, 11 Mira Prospekt, Omsk, 644050, Russian Federation, PhD student. E-mail: yogorodnikov@gmail.com

This paper is devoted to modification of the continuous method of finding solution to the satisfiability of Boolean formulas (SAT). It shows how the method by which the functional associated with SAT is built and the algorithm of its solution by the method of simple iteration is proposed. There are references to the author's previous works and it is explained that there is no way to find a strict solution because the trajectory of the methods leads to a ravine. Instead

[*] *Received 26 September 2014.*

of constructing heuristics to overcome the ravine situation and continue the search it is proposed to construct a method that will identify some bits approximation with high probability. The results can be successfully used in problems of cryptanalysis of asymmetric ciphers, logistics, automatic testing, data recognition, authorization, in various problems on graphs and other problems which are reduced to the SAT problem by a polynomial algorithm. Particular attention is paid to the integer factorization problem on which a well-known RSA asymmetric encryption algorithm is built. It is known that a part of a SAT satisfying set is a RSA encryption key. So, it is sufficient to be able to identify a portion of a satisfying set with high probability. In previous papers some zero bits were identified with high probability. In this paper the emphasis is on the correct identification of individual bits. To increase the number of single bits a nontrivial projection of real variables to Boolean ones is constructed. Four ways of projection are suggested: projection after all iterations, projection between iterations with fixed parameters, projection with different parameters, and the Bayesian approach. The testing of methods is performed on 3-CNF instances equivalent to the integer factorization problem. The results are compared with each other, and the optimality criterion performs a certain number of true bits (both zero and single). Random quantities ξ_0 and ξ_1 equal to the number of correct determination of zero and one bits respectively are introduced. Using the Pearson chi-square method it is shown that ξ_0 and ξ_1 have a uniform distribution, which allows choosing an arbitrary starting approximation.

Keywords: method of simple iteration, satisfiability of Boolean formulas, projection of real variables to Boolean, parameter of projection, Bayesian approach, number of true-determined bits, uniform distribution, single bits

DOI: 10.17212/1814-1196-2015-1-183-200

REFERENCES

1. Levin L.A. Universal'nye zadachi perebora [Universal search problems]. *Problemy peredachi informatsii – Problems of Information Transmission*, 1973, vol. 9, iss. 3, pp. 265–266. Translated from *Problemy peredachi informatsii*, 1973, vol. 9, iss. 3, pp. 115–116.

2. Sat Live! Available at: www.satlive.org (accessed 28.01.2015).

3. Gu J. Local search for satisfiability (SAT) problem. *IEEE Transactions on Systems, Man, and Cybernetics*, 1993, vol. 23, iss. 4, pp. 1108–1129. doi: 10.1109/21.247892

4. Gu J. On optimizing a search problem. *Artificial Intelligence Methods and Applications*. N.G. Bourbakis, ed. New Jersey, World Scientific Publishing, 1992, vol. 1, ch. 2, pp. 63–105.

5. Cook S.A. The complexity of theorem proving procedures. *Proceedings of the Third Annual Symposium on Theory of Computing*. New York, USA, ACM Press, 1971, pp. 151–158. doi: 10.1145/800157.8050476

6. Davis M., Logeman G., Loveland D. A machine program for theorem proving. *Communication of the ACM*, 1962, vol. 5, no. 7, pp. 394–397. doi: 10.1145/368273.368557

7. Batiti R., Protasi M. Approximate Algorithms and Heuristics for MAX-SAT. *Handbook of Combinatorial Optimization*. Dordrecht, Kluwer Academic Publishers, 1998, vol. 1, pp. 77–148. doi: 10.1007/978-1-4613-0303-9_2

8. Guseletova O.N. *Matematicheskie modeli i algoritmy diskretnoi optimizatsii dlya resheniya zadach formirovaniya slozhnykh izdelii*. Avtoref. diss. kand. tekhn. nauk. [Mathematical models and algorithms for discrete optimization for solving the odds-ming of complex products. Author's abstract of PhD eng. sci. diss.]. Omsk, 2008. 16 p.

9. Menezes A.J., Oorschot P.C. van, Vanstone S.A. *Handbook of applied cryptography*. Boca Raton, Florida, USA, CRC Press, 2001. 780 p.

10. Maslov S.Yu. [Iterative methods for linear search in the model as a model of intuitive]. *Tezisy IX Vsesoyuznogo simpoziuma po kibernetike*, 10–15 noyabrya 1981 g. [Abstracts of the All-Union-IX Symposium on Cybernetics], Sukhumi, 10–15 November 1981, pp. 26–28.

11. Kreinovich V.Ya. [The semantics of an iterative method S.Yu. Maslov]. *Sbornik statei Nauchnogo soveta po kompleksnoi probleme "Kibernetika" AN SSSR "Voprosy kibernetiki. Problemy*

sokrashcheniya perebora [Collected papers of the Scientific Council on the complex problem "Cybernetics", USSR Academy of Sciences "Problems of Cybernetics. Problems with the reduction-busting"]. Moscow, 1987, iss. 131, pp. 30–62.

12. Oparin G.A., Novopashin A.P. [Continuous models of solutions of systems of Boolean equations]. *Vestnik Tomskogo gosudarstvennogo universiteta. Prilozhenie: Materialy mezhdunarodnykh vsesoyuznykh i regional'nykh nauchnykh konferentsii, simpoziumov, shkol, prokhodivsh – Tomsk State University Journal. Supplement: Proceedings of the International Union and regional scientific conferences, workshops, schools, held at TSU*, 2004, no. 9 (1), pp. 20–25.

13. Faizullin R.T., Dul'keit V.I., Ogorodnikov Yu.Yu. Gibridnyi metod poiska priblizhennogo resheniya zadachi 3-vypolnimost', assotsiirovannoi s zadachei faktorizatsii [Hybrid method for the approximate solution of the 3-satisfiability problem associated with the factorization problem]. *Trudy Instituta matematiki i mekhaniki UrO RAN – Proceedings of the Steklov Institute of Mathematics. Supplement. Proceedings of the Institute of Mathematics and Mechanics Ural Branch of RAS*, 2013, vol. 19, iss. 2, pp. 285–294.

14. Dul'keit V.I., Faizullin R.T., Khnykin I.G. Nepreryvnye approksimatsii resheniya zadachi «vypolnimost"» primenitel'no k kriptograficheskomu analizu asimmetrichnykh shifrov [Continuous approximation of SAT decision as applied to cryptographic analysis of asymmetric ciphers]. *Komp'yuternaya optika – Computer optics*, 2009, vol. 33, no. 1, pp. 68–73.

15. Ogorodnikov Yu.Yu., Faizullin R.T. Opredelenie nulevykh bit zadachi 3-vypolnimost', assotsiirovannoi s zadachei faktorizatsii [Recognition of zero bits of 3-SAT problem by applying linear algebra's methods]. *Komp'yuternaya optika – Computer optics*, 2014, vol. 38, no. 3, pp. 521–528.

16. Rivest R.L., Shamir A., Adleman L. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 1978, vol. 21, iss. 2, pp. 120–126. doi: 10.1145/359340.359342

17. Patsakis C. RSA private key reconstruction from random bits using SAT solvers. Cryptology ePrint Archive: report 2013/026. Received 18.01.2013, last revised 7.05.2013. Available at: https://eprint.iacr.org/2013/026 (accessed 28.01.2015)

18. Berezin I.S., Zhidkov N.P. *Metody vychislenii. T. 2. Glava 7.5. Reshenie sistem uravnenii* [Computing methods. Vol. 2, ch. 7.5. Solving systems of equations]. Moscow, Fizmatlit Publ., 1959, pp. 150–162.

19. Mestetskii L.M. *Matematicheskie metody raspoznavaniya obrazov. Gl. 2.1. Klassifikatsiya na osnove Baiesovskoi teorii reshenii* [Mathematical methods of pattern recognition. Ch. 2.1. Classification based on Bayesian decision theory]. Moscow, Department of Mathematical Methods of Forecasting Faculty of Computational Mathematics and Cybernetics of Lomonosov Moscow State University, 2002, pp. 8–13. Available at: http://www.ccas.ru/frc/papers/mestetskii04course.pdf (accessed 28.01.2015)

20. Kibzun A.I., Goriyanova A.R., Naumov A.V., Sirotin A.N. *Teoriya veroyatnostei i matematicheskaya statistika. Bazovyi kurs s primerami i zadachami* [The theory of probability and mathematical statistics. Basic course with examples and problems]. Moscow, Fizmatlit Publ., 2002. 224 p.