2012 Вычислительные методы в дискретной математике №3(17)

DOI 10.17223/20710410/17/11

УДК 510.53+519.61

О НЕКОТОРЫХ СВОЙСТВАХ ОБРАЗОВ ТРАНСФОРМИРОВАННЫХ ЗАДАЧ

Д. М. Мурин

Ярославский государственный университет им. П. Г. Демидова, г. Ярославль, Россия

E-mail: nirum87@mail.ru

Одним из способов доказательства NP-полноты задачи является полиномиальное сведение (трансформация) к ней задачи, NP-полнота которой уже установлена. При этом изучению свойств полученного образа, на наш взгляд, уделяется недостаточно внимания. В 1985 г. в работе Дж. Лагариаса и А. М. Одлыжко предложен метод решения NP-полной задачи о рюкзаке, дающий верное решение для «практически всех» рюкзаков, плотность которых не превышает 0,6463... В настоящей работе рассматривается вопрос о том, в какую область задач о рюкзаке (относительно плотности рюкзаков) при доказательстве NP-полноты попадают образы следующих задач: 3-ВЫП, Раскрашиваемость, Точное покрытие.

Ключевые слова: *NP-полнота, метод Лагариаса — Одлыжко, задача о рюкзаке.*

Введение

«Задача» и «алгоритм» являются центральными понятиями теории сложности вычислений. Успехи, достигнутые за последние десятилетия в этой теории, в значительной степени связаны с рассмотрением понятия алгоритма в качестве первостепенного, что отразилось в несколько изменённом наименовании самой теории — теория сложности алгоритмов.

На наш взгляд, в теории сложности вычислений правомерны и должны быть интересны вопросы о «внутренней» структуре задачи и о структуре задачи в соотнесении с другими задачами.

Исходя из понимания задачи (массовой задачи) как множества (как правило, бесконечной серии) индивидуальных задач [1], при изучении структуры задач нам кажется естественным использовать язык теории множеств. Так, вместо полиномиальной сводимости (трансформации) задачи P_1 к задаче P_2 будем говорить о полиномиальном вложении задачи P_1 в задачу P_2.

Напомним формулировки рассматриваемых в работе задач.

3-ВЫП.

Условие: Булева формула F в 3-КНФ.

Вопрос: Выполнима ли булева формула F?

Раскрашиваемость.

Условие: Неориентированный граф $G = (V, E)$, натуральное число k.

Вопрос: Является ли граф G k-раскрашиваемым?

Точное покрытие.

Условие: Семейство подмножеств $M = \{M_1, \ldots, M_l\}$ множества $\mathcal{U} = \{u_1, \ldots, u_p\}$.

Вопрос: Существует ли для семейства M подсемейство попарно непересекающихся множеств, являющееся покрытием множества $\mathcal{U}$?

Задача о рюкзаке.

Условие: $(a_1, a_2, \ldots, a_r, b) \in \mathbb{N}^{r+1}$.

Вопрос: Имеет ли уравнение $\sum\limits_{j=1}^{r} a_j x_j = b$, *где* x_j, $j = 1, \ldots, r$, — *неизвестные, решение в числах* 0 *и* 1*?*

В работе [2] рассматривается понятие *плотности* рюкзака.

Определение 1. *Плотностью* рюкзака $(a_1, a_2, \ldots, a_r, b)$ называется число

$$d = \frac{r}{\log_2(\max\limits_{1 \leqslant i \leqslant r} a_i)}.$$

Рассмотрим задачу о рюкзаке $(a_1, a_2, \ldots, a_r, s) \in \mathbb{N}^{r+1}$, где s — сумма элементов произвольного подмножества множества $\{a_1, a_2, \ldots, a_r\}$; иными словами, $s = \sum\limits_{i=1}^{r} e_i a_i$, где $e_i \in \{0, 1\}, i = 1, \ldots, r$. Только для задач этого типа ответом на вопрос задачи о рюкзаке является «Да».

Пусть $t = \sum\limits_{i=1}^{r} a_i$, тогда можно считать, что $s \geqslant t/n$; иначе, если $s < t/n$, то существует $a_{i_0} \geqslant t/n$ и вопрос «Имеет ли уравнение $\sum\limits_{j=1}^{r} a_j x_j = b$, где x_j, $j = 1, \ldots, r$, — неизвестные, решение в числах 0 и 1?» можно заменить вопросом «Имеет ли уравнение $\sum\limits_{j=1, j \neq i_0}^{r} a_j x_j = b$, где x_j, $j = 1, \ldots, \hat{i_0}, \ldots, r$, — неизвестные, решение в числах 0 и 1?», уменьшив тем самым размерность задачи о рюкзаке. Аналогично, $s \leqslant (n-1)t/n$.

Определим следующим образом векторы $b_1, \ldots, b_{r+1}$ размерности $r + 1$:

$$\begin{gathered} b_1 = (1, 0, \ldots, 0, N \cdot a_1), \\ b_2 = (0, 1, \ldots, 0, N \cdot a_2), \\ \ldots \\ b_r = (0, 0, \ldots, 1, N \cdot a_r), \\ b_{r+1} = (0, 0, \ldots, 0, N \cdot s), \end{gathered}$$

где $N \geqslant \lceil \sqrt{r/2} \rceil$ — некоторое достаточно большое натуральное число. Очевидно, векторы $b_1, \ldots, b_{r+1}$ линейно независимы. Пусть Λ — решётка, образованная этими векторами: $\Lambda = \{z_1 b_1 + \cdots + z_{r+1} b_{r+1} : z_1, \ldots, z_{r+1} \in \mathbb{Z}\}$.

Предположим, что у нас имеется АЛГОРИТМ $\mathfrak{A}$, который за полиномиальное время выдает один из кратчайших ненулевых векторов решётки. Ситуация с данным АЛГОРИТМом $\mathfrak{A}$ следующая: задача поиска кратчайшего ненулевого вектора решётки NP-полна, однако LLL-алгоритм построения приведённого базиса решетки позволяет за полиномиальное время[1] получать «достаточно короткие» векторы решётки [3]. В 1985 г. Дж. Лагариас и А. М. Одлыжко [2] установили, что при плотности рюкзака меньше 0,6463... с ростом размерности рюкзака вероятность того, что произвольный кратчайший вектор решётки Λ не даёт решения задачи о рюкзаке, стремится к нулю. Этот результат улучшен в работе [4] до плотности рюкзака меньше 0,9408... благодаря следующей модификации рассматриваемой решётки Λ: векторы $b_1, \ldots, b_r$ остаются прежними, $b_{r+1} = (1/2, 1/2, \ldots, 1/2, N \cdot s)$, $N \geqslant \lceil \sqrt{r}/2 \rceil$.

Далее в работе излагается доказательство следующей теоремы.

[1] Сложность LLL-алгоритма построения приведённого базиса решётки Λ составляет $O((r+1)^6 \log_2^3 B)$ битовых операций, где $B \geqslant |b_i|^2$ для $1 \leqslant i \leqslant r+1$ [3].

Теорема 1. Существует последовательность полиномиальных трансформаций (полиномиальных вложений) задачи 3-ВЫП в задачу о рюкзаке (3-ВЫП $\propto$ Раскрашиваемость $\propto$ Точное покрытие $\propto$ Задача о рюкзаке), обладающая тем свойством, что образ задачи 3-ВЫП лежит в области задач о рюкзаке, с высокой вероятностью решаемых методом Лагариаса — Одлыжко.

1. Точное покрытие $\propto$ Задача о рюкзаке

В работе Р. М. Карпа [5] задача о точном покрытии следующим образом полиномиально вкладывается в задачу о рюкзаке.

Пусть $w = |M| + 1 = l + 1$, а для всех $1 \leqslant j \leqslant |M|$ и $1 \leqslant i \leqslant p$

$$\varepsilon_{ji} = \begin{cases} 1, & \text{если } u_i \in M_j, \\ 0, & \text{если } u_i \notin M_j. \end{cases}$$

Тогда полагаем $r = |M|$, $a_j = \sum\limits_{i=1}^{p} \varepsilon_{ji} w^{i-1}$ для всех $1 \leqslant j \leqslant |M|$ и $b = (w^p - 1)/(w - 1)$. Таким образом, задача о рюкзаке приобретает следующий вид:

$$\left(\sum_{i=1}^{p} \varepsilon_{1i} w^{i-1}, \ldots, \sum_{i=1}^{p} \varepsilon_{ri} w^{i-1}, \frac{w^p - 1}{w - 1} \right).$$

Отметим, что вопрос о вхождении u_p в одно из множеств семейства M может быть решен за $\mathrm{O}(lp)$ арифметических операций. Если u_p не входит ни в одно множество семейства M, то для семейства M подсемейства попарно непересекающихся множеств, являющегося покрытием множества $\mathcal{U}$, не существует.

Оценим плотность получаемых описанным выше образом рюкзаков в предположении, что u_p входит в те множества семейства M, чьи индексы содержатся в множестве I. Для всех $1 \leqslant j \leqslant |M|$ справедлива цепочка неравенств

$$\sum_{i=1}^{p-1} \varepsilon_{ji} w^{i-1} \leqslant \sum_{i=1}^{p-1} w^{i-1} = (w^{p-1} - 1)/(w - 1) < w^{p-1},$$

поэтому

$$\max_{1 \leqslant j \leqslant |M|} \sum_{i=1}^{p} \varepsilon_{ji} w^{i-1} = \max_{j \in I} \sum_{i=1}^{p} \varepsilon_{ji} w^{i-1} = w^{p-1} + \max_{j \in I} \sum_{i=1}^{p-1} \varepsilon_{ji} w^{i-1},$$

из чего следует оценка

$$d_{\text{ТП}} = \frac{|M|}{\log_2 \left(\max\limits_{1 \leqslant j \leqslant |M|} \sum\limits_{i=1}^{p} \varepsilon_{ji} w^{i-1} \right)} = \frac{|M|}{\log_2 \left(w^{p-1} + \max\limits_{j \in I} \sum\limits_{i=1}^{p-1} \varepsilon_{ji} w^{i-1} \right)} \leqslant$$
$$\leqslant \frac{|M|}{\log_2(w^{p-1})} = \frac{|M|}{(p-1)\log_2 w} = \frac{|M|}{(p-1)\log_2(|M|+1)}.$$

Таким образом, при полиномиальном вложении задачи о точном покрытии в задачу о рюкзаке образ задач о точном покрытии, удовлетворяющих условию $\dfrac{|M|}{(p-1)\log_2(|M|+1)} \leqslant 0{,}9408\ldots$, лежит в области задач о рюкзаке, с высокой вероятностью решаемых методом Лагариаса — Одлыжко.

Проиллюстрируем поведение функции

$$F(p, l) = \frac{l}{(p-1)\log_2(l+1)}.$$

На рис. 1 показаны плоскость 0,9408... и значения функции $F(p,l)$ при $2 \leqslant p \leqslant 100$, $1 \leqslant l \leqslant 300$ (p и l — натуральные числа). Из рисунка видно, что задачи о точном покрытии, удовлетворяющие условию $\dfrac{|M|}{(p-1)\log_2(|M|+1)} \leqslant 0{,}9408...$, составляют значительную долю задач о точном покрытии. Оценим эту долю.

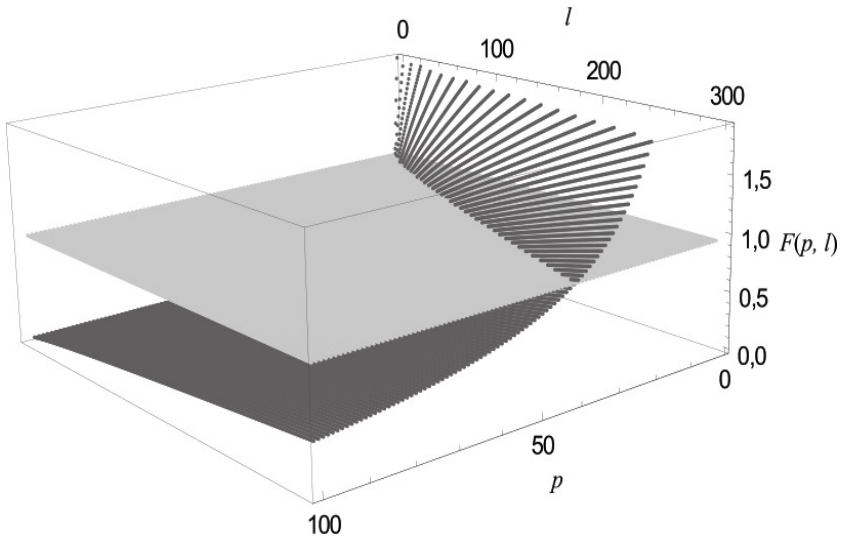

Рис. 1. Поведение функции $F(p,l)$ при $2 \leqslant p \leqslant 100$, $1 \leqslant l \leqslant 300$

Кривая $p = C_1 l/(\log_2(l+1)) + 1$, где $C_1 = (0{,}9408...)^{-1}$, делит положительный квадрант плоскости 0,9408... на две области (рис. 2). Область над кривой соответствует значениям функции $F(p,l)$, меньшим 0,9408... (задачи о точном покрытии с параметрами p, l из данной области с высокой вероятностью решаются методом Лагариаса — Одлыжко), назовём эту область «хорошей». Область под кривой соответствует значениям функции $F(p,l)$, большим 0,9408..., назовем эту область «плохой».

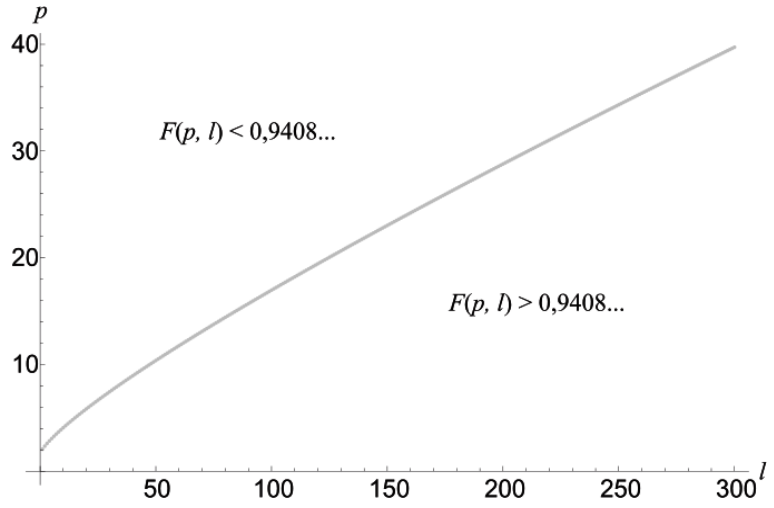

Рис. 2. Разделение положительного квадранта плоскости 0,9408... кривой $p = \frac{C_1 l}{\log_2(l+1)} + 1$

Площадь «плохой» области на интервале $1 \leqslant l \leqslant x-1$ может быть вычислена следующим образом:

$$P_{\text{по}}(x-1) = \int\limits_1^{x-1} \left(\frac{C_1 l}{\log_2(l+1)} + 1 \right) dl = x - 2 + \int\limits_1^{x-1} \frac{C_2(l+1)}{\ln(l+1)} d(l+1) - C_2 \int\limits_1^{x-1} \frac{d(l+1)}{\ln(l+1)} =$$

$$= x - 2 + C_2 \int\limits_1^{x-1} \frac{d(l+1)^2}{\ln(l+1)^2} - C_2 \mathrm{Li}(x) = x - 2 + C_2 \mathrm{Li}(x^2) - C_2 \mathrm{Li}(x),$$

где $C_2 = C_1 / \log_2 e$.

Таким образом, доля «плохой» области в квадрате, ограниченном прямыми $p = x-1$, $l = x-1$ и $p = 1$, $l = 1$, составляет

$$\frac{P_{\text{по}}(x-1)}{(x-2)^2} = \frac{x - 2 + C_2 \mathrm{Li}(x^2) - C_2 \mathrm{Li}(x)}{(x-2)^2}.$$

Воспользуемся тем фактом [6], что при $x \to \infty$

$$\mathrm{Li}(x) = x \left(\sum_{i=1}^{n} \frac{(i-1)!}{(\ln x)^i} + \mathrm{O}\left(\frac{1}{(\ln x)^{n+1}} \right) \right),$$

и перейдём к пределу при $x \to \infty$. Получим

$$\lim_{x\to\infty} \frac{P_{\text{по}}(x-1)}{(x-2)^2} = \lim_{x\to\infty} \frac{x - 2 + C_2 \mathrm{Li}(x^2) - C_2 \mathrm{Li}(x)}{(x-2)^2} =$$

$$= \lim_{x\to\infty} \frac{1}{x-2} + \lim_{x\to\infty} C_2 \left(\sum_{i=1}^{n} \frac{x^2 (i-1)!}{(x-2)^2 (\ln x^2)^i} + \mathrm{O}\left(\frac{x^2}{(x-2)^2 (\ln x)^{n+1}} \right) \right) -$$

$$- \lim_{x\to\infty} C_2 \left(\sum_{i=1}^{n} \frac{x (i-1)!}{(x-2)^2 (\ln x)^i} + \mathrm{O}\left(\frac{x}{(x-2)^2 (\ln x)^{n+1}} \right) \right) = 0.$$

Данный результат подтверждает тот факт, что доля «плохих» параметров p и l мала по сравнению с общим числом параметров задачи о точном покрытии.

Так как у каждого подмножества множества $\mathcal{U}$ есть дополнение до $\mathcal{U}$ и ввиду нецелесообразности включения в семейство M пустого множества и лёгкости задачи о точном покрытии в случае, если $\mathcal{U} \in M$, можно считать, что в семейство M входят не более чем $2^{p-1} - 1$ множеств. Иными словами, $l \leqslant 2^{p-1} - 1$, или $\log_2(l+1) \leqslant p - 1$.

Рассмотрим два интересных, на наш взгляд, случая.

1. Если $l + 1 = 2^{\frac{p-1}{C_3}}$, где C_3 — некоторая константа, то число удовлетворяющих критерию $F(p,l) \leqslant 0{,}9408...$ параметров p и l конечно и зависит только от C_3. Действительно, в этом случае

$$\frac{l}{(p-1)\log_2(l+1)} = \frac{l}{C_3 \log_2^2(l+1)},$$

и число l, таких, что $l \leqslant C_4 \log_2^2(l+1)$, где $C_4 = C_3 \cdot 0{,}9408...$, зависит только от C_3.

2. Если $l = C_5(p-1)$, где C_5 — некоторая константа, то число удовлетворяющих критерию $F(p,l) \leqslant 0{,}9408...$ параметров p и l бесконечно велико, так как неравенству $C_5 / \log_2(l+1) \leqslant 0{,}9408...$ удовлетворяет бесконечное число параметров l.

2. Раскрашиваемость $\propto$ Точное покрытие

В работе [7] задача раскрашиваемости следующим образом полиномиально вкладывается в задачу о точном покрытии.

Пусть $G = (V, E)$ — неориентированный граф и k — натуральное число, определяющие задачу раскрашиваемости.

Тогда задача о точном покрытии строится следующим образом:

$$\mathcal{U} = V \cup \{[e, i] : e \in E \text{ и } 1 \leqslant i \leqslant k\} \text{ и } M = \{S_{vi}\} \cup \{T_{ei}\},$$

где $S_{vi} = v \cup \{[e, i] :$ ребро e инцидентно $v\}$ для каждого узла $v \in V$ и $1 \leqslant i \leqslant k$ и $T_{ei} = \{[e, i]\}$ для каждого ребра $e \in E$ и $1 \leqslant i \leqslant k$.

Плотность d_{P} рюкзаков, получаемых путём последовательных полиномиальных вложений задачи о раскрашиваемости в задачу о точном покрытии, а затем задачи о точном покрытии в задачу о рюкзаке, может быть оценена следующим образом:

$$d_{\text{P}} \leqslant \frac{|M|}{(|\mathcal{U}| - 1)\log_2(|M| + 1)} = \frac{|\{S_{vi}\} \cup \{T_{ei}\}|}{(|V| + |E|k - 1)\log_2(|\{S_{vi}\} \cup \{T_{ei}\}| + 1)} =$$
$$= \frac{|V|k + |E|k}{(|V| + |E|k - 1)\log_2(|V|k + |E|k + 1)} = \frac{(|V| + |E|)k}{(|V| + |E|k - 1)\log_2((|V| + |E|)k + 1)}.$$

Таким образом, при полиномиальном вложении задачи раскрашиваемости в задачу о рюкзаке образ задач раскрашиваемости, удовлетворяющих условию

$$\frac{(|V| + |E|)k}{(|V| + |E|k - 1)\log_2((|V| + |E|)k + 1)} \leqslant 0{,}9408...,$$

лежит в области задач о рюкзаке, с высокой вероятностью решаемых методом Лагариаса — Одлыжко.

3. 3-ВЫП $\propto$ Раскрашиваемость

В [7] задача 3-ВЫП следующим образом полиномиально вкладывается в задачу раскрашиваемости.

Пусть дана формула в 3-КНФ с n переменными и t сомножителями, $x_1, x_2, \ldots, x_n$ и $F_1, F_2, \ldots, F_t$ — соответственно переменные и сомножители формулы F, $v_1, v_2, \ldots, v_n$ — новые символы. Тогда полагаем $k = n + 1$, узлы графа G таковы:

1) $x_i, \overline{x_i}, v_i$ для $1 \leqslant i \leqslant n$;
2) F_i для $1 \leqslant i \leqslant t$.

Рёбра графа G:

1) все (v_i, v_j), для которых $i \neq j$;
2) все (v_i, x_j) и $(v_i, \overline{x_j})$, для которых $i \neq j$;
3) $(x_i, \overline{x_i})$ для $1 \leqslant i \leqslant n$;
4) (x_i, F_j), если x_i не входит в F_j, и $(\overline{x_i}, F_j)$, если $\overline{x_i}$ не входит в F_j.

Таким образом, $|V| = 3n + t$, $|E| = n(n-1)/2 + n(n-1) + n + (2n-3)t$, т. е.

$$|E| = 1{,}5n^2 - 0{,}5n + (2n - 3)t,$$
$$|V| + |E| = 1{,}5n^2 + 2{,}5n + 2(n - 1)t,$$
$$[|V| + |E|(n + 1) = 1{,}5n^3 + n^2 + 2{,}5n + ((2n - 1)n - 2)t.$$

Предполагается, что в одном сомножителе дважды одна переменная или её отрицание не встречаются.

Плотность $d_{3\text{-ВЫП}}$ рюкзаков, получаемых путём последовательных трансформаций задачи 3-ВЫП в задачу о рюкзаке, может быть оценена следующим образом:

$$d_{3\text{-ВЫП}} \leqslant \frac{(|V|+|E|)k}{(|V|+|E|k-1)\log_2((|V|+|E|)k+1)} <$$
$$< \frac{(1{,}5n^2+2{,}5n+2(n-1)t)(n+1)}{(1{,}5n^3+n^2+2{,}5n+(2n^2-n-2)t-1)\log_2((1{,}5n^2+2{,}5n+2(n-1)t)(n+1))} =$$
$$= \frac{3n^3+8n^2+5n+4(n^2-1)t}{(3n^3+2n^2+5n+2(2n^2-n-2)t-2)\log_2(1{,}5n^3+4n^2+2{,}5n+2(n^2-1)t)}.$$

Так как $n^3 > n$ при $n > 1$, $n^2-n-1 > 0$ при $n > (1+\sqrt{5})/2$ и $3n^3+2n^2+5n-2 > 2n^3+4n^2+2n$ при $n > 1$, то

$$\frac{3n^3+8n^2+5n+4(n^2-1)t}{(3n^3+2n^2+5n+2(2n^2-n-2)t-2)\log_2(1{,}5n^3+4n^2+2{,}5n+2(n^2-1)t)} <$$
$$< \frac{4n^3+8n^2+4n+4(n^2-1)t}{(2n^3+4n^2+2n+2(n^2-1)t)\log_2(1{,}5n^3+4n^2+2{,}5n+2(n^2-1)t)} =$$
$$= \frac{2}{\log_2(1{,}5n^3+4n^2+2{,}5n+2(n^2-1)t)} < \frac{2}{\log_2(2(n+1)t)}.$$

Неравенство $2/\log_2(2(n+1)t) \leqslant 0{,}9408...$ выполняется при $(n+1)t \geqslant 2^{1{,}1258...}$, что, очевидно, выполнено при $n \geqslant 2$ и $t \geqslant 1$, случай же $n = 1$ выглядит с точки зрения задачи 3-ВЫП очень экзотично.

Таким образом, при полиномиальном вложении задачи 3-ВЫП в задачу о рюкзаке образ всех заслуживающих внимания задач 3-выполнимости лежит в области задач о рюкзаке, с высокой вероятностью решаемых методом Лагариаса — Одлыжко.

Теорема доказана.

На наш взгляд, применимость метода Лагариаса — Одлыжко к решению на практике NP-полных задач является интересной темой для экспериментальных исследований. Разработанный в целях проведения этих исследований программный комплекс проходит процедуру тестирования.

ЛИТЕРАТУРА

1. *Успенский В. А., Семенов А. Л.* Теория алгоритмов: основные открытия и приложения. М.: Наука, 1987. 288 с.
2. *Odlyzko A. M. and Lagarias J. C.* Solving Low-Density Subset Sum Problems // J. Association Computing Machinery. 1985. V. 32. No. 1. P. 229–246.
3. *Василенко О. Н.* Теоретико-числовые алгоритмы в криптографии. М.: МЦНМО, 2006. 336 с.
4. *Coster M. J., Joux A., LaMacchia B. A., et al.* Improved low-density subset sum algorithms // Computational Complexity. 1992. No. 2. P. 111–128.
5. *Karp R. M.* Reducibility among combinatorial problems // Complexity of Computer Computations: Proc. of a Symp. on the Complexity of Computer Computations, the IBM Research Symposia Series. NY: Plenum Press, 1972. P. 85–103.
6. *Коксон Э. Т.* Асимптотические разложения. М.: Мир, 1966. 159 с.
7. *Ахо А., Хопкрофт Дж., Ульман Дж.* Построение и анализ вычислительных алгоритмов. М.: Мир, 1979. 536 с.