Математика и ее приложения : ЖИМО. 2012. Вып. 1 (9). С. 75–82

УДК 512.54

А. А. Толстопятов[1]

Алгоритм получения алгебраических зависимостей между булевыми полиномами

Ключевые слова: булево сжатие, булевы полиномы.

Построен алгоритм получения всех алгебраических зависимостей между булевыми полиномами. Получена оценка числа этих зависимостей. Рассмотрен частный случай двух булевых переменных.

Key words: boolean compress, boolean polynoms.

An algorithm of the receipt of all algebraic dependencies between Boolean polynomials. The estimate of the number of these dependencies. Considered a special case of the two Boolean variables.

Булево сжатие основано на поиске зависимостей между булевыми полиномами, кодирующими первое поле принадлежности буферов [2]. Главной проблемой при таком подходе к сжатию является построение алгоритма разбиения файла на буферы, чтобы указанные выше булевы полиномы удовлетворяли кодирующему уравнению [3]. Пока единственный подход к решению этой задачи, доведенный до алгоритма, предложенный в [1] носит академический характер, а не прикладной, так как он основан на использовании экспоненциально большого множества исходных порождающих и получен при двух ограничениях. Во-первых, порождающими служит часть булевых полиномов, кодирующих поля принадлежности. Во-вторых, предположим, что эти полиномы таковы, что через них явно можно выразить все n булевых переменных. При построении алгоритмов, свободных от этих ограничений, возникает задача получения всех зависимостей между произвольным множеством булевых полиномов и оценки числа таких зависимостей. Решению этой задачи и посвящена настоящая работа.

Пусть поля принадлежности L буферов, закодированных булевыми полиномами $f_l(x_i), i = 1, \ldots, n; l = 1, \ldots, L$. Разложим эти полиномы по полиномам Лагранжа:

$$f_l(x_i) = \sum_{j=0}^{2^n-1} C_{lj} L_j(x_i), \tag{1}$$

где

$$L_j(x_i) = \prod_{k=1}^{n} L_{j_k}(x_k) \tag{2}$$

[1]Ивановский государственный университет; E-mail: khash2@mail.ru. Работа выполнена при финансовой поддержке РФФИ (проект 10-07-00350а)

а полиномы Лагранжа от одной булевой переменной x_i есть

$$L_0(x_i) = x_i + 1; \quad L_1(x_i) = x_i. \tag{3}$$

Числа j_k из $GF(2)$ есть коэффициенты в двоичном коде натурального числа j:

$$j = \sum_{k=1}^{n} j_k 2^{k-1}. \tag{4}$$

Полиномы Лагранжа L_j удовлетворяют условию

$$L_i L_j = \delta_{ij} L_i. \tag{5}$$

Так как булевы полиномы f_l удовлетворяют условиям:

$$f_l^2 = f_l; \; f_l + f_l = 0, \tag{6}$$

то их можно рассматривать в качестве булевых переменных. От этих переменных можно построить полиномы Лагранжа $L_m(f_l), m = 0, 1, \ldots, 2^L - 1$, которые есть:

$$L_m(f_l) = \prod_{k=1}^{L} L_{m_k}(f_k), \tag{7}$$

где m_k определяется из:

$$m = \sum_{k=1}^{L} m_k 2^{k-1}. \tag{8}$$

Используя определение полиномов Лагранжа от одной переменной (3), можно переписать (7) так:

$$L_m(f_l) = \prod_{k=1}^{L} (j_m + 1 + f_m) = \prod_{k=1}^{L} (j_m + 1 + \sum_{j=0}^{2^n-1} C_{mj} L_j(x_i)) \; . \tag{9}$$

Так как разложение булевых полиномов $f \equiv 1$ и $f \equiv 0$ по полиномам Лагранжа имеет вид:

$$1 = \sum_{j=0}^{2^n-1} 1 \cdot L_j(x_i); \; 0 = \sum_{j=0}^{2^n-1} 0 \cdot L_j(x_i), \tag{10}$$

то (9) можно переписать так:

$$L_m(f_l) = \prod_{k=1}^{L} (j_m + \sum_{j=0}^{2^n-1} (C_{m_j} + 1) L_j(x_i)) \; . \tag{11}$$

Из (7) и (8) следует, что (11) представимо в виде:

$$L_m(f_l) = \prod_{k=1}^{L} (j_m + 1 + f_m) = \prod_{k=1}^{L} (j_m + 1 + \sum_{j=0}^{2^n-1} C_{m_j} L_j(x_i)), \tag{12}$$

так как для $m_k \in \{0,1\}$ выполнены:

$$\begin{aligned} m_k = 1 \quad &= \sum_{j=0}^{2^n-1} 1 \cdot L_j(x_i), \\ m_k = 0 \quad &= \sum_{j=0}^{2^n-1} 0 \cdot L_j(x_i), \end{aligned} \tag{13}$$

то (12) представимо в виде:

$$L_m(f_l) = \prod_{k=1}^{L}(j_m + 1 + \sum_{j=0}^{2^n-1}(C_{m_j} + 1)L_j(x_i)), \tag{14}$$

и далее:

$$L_m(f_l) = \prod_{k=1}^{L}(j_m + 1 + \sum_{j=0}^{2^n-1}(m_k + C_{m_j} + 1)L_j(x_i)). \tag{15}$$

Формула (15) дает разложение полиномов Лагранжа $L_m(f_l)$ по полиномам Лагранжа $L_j(x_i)$.

Правая часть любой зависимости между булевыми полиномами f_l, которую обозначим через $F(f_l) = 0$, может быть разложена по полиномам Лагранжа $L_m(f_l)$:

$$F(x_l) = \sum_{m=0}^{2^L-1} a_m L_m(f_l), \tag{16}$$

или, используя (15), – по полиномам Лагранжа $L_j(x_i)$:

$$F(x_l) = \sum_{m=0}^{2^L-1} a_m \prod_{k=1}^{L} \sum_{j=0}^{2^n-1}(m_k + C_{mj} + 1)L_j(x_i). \tag{17}$$

В силу (5) в (17) можно поменять местами сумму и произведение. Тогда получим:

$$F(x_l) = \sum_{m=0}^{2^L-1} a_m \sum_{j=0}^{2^n-1} \left[\prod_{k=1}^{L}(m_k + C_{mj} + 1)\right] L_j(x_i). \tag{18}$$

Переставляя в (18) суммы по j и по m и приравнивая $F(x_l)$ к нулю, будем иметь:

$$F(x_l) = \sum_{j=0}^{2^n-1} \left[\sum_{m=0}^{2^L-1} a_m \prod_{k=1}^{L}(m_k + C_{mj} + 1)\right] L_j(x_i) = 0. \tag{19}$$

Из (5) следует, что полиномы Лагранжа $L_j(x_i)$ линейно независимы. А значит из (19) получим:

$$\sum_{m=0}^{2^L-1} a_m \prod_{k=1}^{L}(m_k + C_{mj} + 1) = 0. \tag{20}$$

Зависимость между булевыми полиномами f_l задается коэффициентами a_m в разложении (16).

Всего таких переменных L. Число булевых уравнений (20) есть 2^n, так как $j = 0, 1, \ldots, 2^n - 1$. Впрочем, любая система булевых уравнений может быть заменена на одно булево уравнение, если от уравнений (20) взять дизъюнкцию:

$$\bigvee_{j=0}^{2^n-1} \left\{ \left[\sum_{m=0}^{2^L-1} a_m \prod_{k=1}^{L} (m_k + C_{mj} + 1) \right] + 1 \right\} + 1 = 0. \tag{21}$$

Левая часть (20) может быть переписана через произведение:

$$\prod_{j=0}^{2^n-1} \left\{ \left[\sum_{m=0}^{2^L-1} a_m \prod_{k=1}^{L} (m_k + C_{mj} + 1) \right] + 1 \right\} + 1 = 0. \tag{22}$$

Уравнение (22) для коэффициентов a_m в разложении (16) полностью решает задачу о построении всех зависимостей между булевыми полиномами f_l.

Число таких зависимостей есть число разных решений уравнения (22). Для оценки этого числа удобно от уравнения (22) вернуться к системе уравнений (20). Если ввести матрицу A_{mj} согласно:

$$A_{mj} = \prod_{k=1}^{L} (m_k + C_{mj} + 1), \tag{23}$$

то система (20) запишется так:

$$\sum_{m=0}^{2^L-1} A_{mj} a_m = 0. \tag{24}$$

Система (24) – линейная и однородная. Это значит, что если обозначить через $\sigma = rang A_{mj}$, то система (24) будет содержать $2^L - 1 - \sigma$ параметров, которым можно придавать значения из $GF(2)$. Тогда, если ξ – число решений (24), то

$$\xi = 2^{2^L - \sigma - 1}. \tag{25}$$

Формула (25) полностью решает задачу о числе зависимостей между произвольным множеством булевых полиномов $f_l(x_i), l = 1, \ldots, L$.

Для иллюстрации полученных результатов рассмотрим простейший частный случай получения зависимостей между каким-нибудь подмножеством многочленов всех булевых полиномов от двух переменных. Если рассмотреть все булевы полиномы от n переменных, то число коэффициентов в каждом из таких полиномов будет равно 2^n. Это значит, что разных полиномов будет 2^{2^n}, а число подмножеств в таком множестве – $2^{2^{2^n}}$. Для $n = 2$ имеем:

$2^2 = 4$ – коэффициенты полинома,

$2^{2^2} = 16$ – число различных уравнений,

$2^{2^{2^2}} = 65536$ – число подмножеств множества булевых полиномов.

Исключим из этого числа $C_{16}^1 = 16$ множеств из одного полинома и $C_{16}^2 = 120$ множеств из двух полиномов, среди которых зависимостей быть не может. В этом нетрудно убедиться, рассмотрев зависимости общего вида:

$$F(f_1, f_2) = a_1 f_1 + a_2 f_2 + a_3 f_1 f_2 = 0. \tag{26}$$

Тогда множеств полиномов, среди которых могут быть зависимые, будет 65400. Рассмотрим простейший случай трех полиномов от двух булевых переменных x_1 и x_2.

Вводя полиномы Лагранжа $L_j(x_i), j = 0, \ldots, 3;\ i = 1, 2$, согласно

$$\begin{array}{ll} 0. & L_0 = (x_1 + 1)(x_2 + 1), \\ 1. & L_1 = (x_1 + 1)x_2, \\ 2. & L_2 = x_1(x_2 + 1), \\ 3. & L_3 = x_1 x_2, \end{array} \tag{27}$$

можем записать три таких полинома f_1, f_2, f_3, разложив их по базису из полиномов Лагранжа $L_j(x_i)$:

$$\begin{array}{ll} f_1 = & a_0 L_0 + a_1 L_1 + a_2 L_2 + a_3 L_3, \\ f_2 = & b_0 L_0 + b_1 L_1 + b_2 L_2 + b_3 L_3, \\ f_3 = & c_0 L_0 + c_1 L_1 + c_2 L_2 + c_3 L_3, \end{array} \tag{28}$$

где $a_0, \ldots, c_3 \in GF(2)$.

Вводя полиномы Лагранжа $L_k(f), k = 0, \ldots, 7$, где в качестве базовых переменных рассматриваются полиномы f_1, f_2, f_3:

$$\begin{array}{lll} 0. & L_0(f) & = (f_1 + 1)(f_2 + 1)(f_3 + 1), \\ 1. & L_1(f) & = (f_1 + 1)(f_2 + 1)(f_3), \\ 2. & L_2(f) & = (f_1 + 1)(f_2)(f_3 + 1), \\ 3. & L_3(f) & = (f_1 + 1)(f_2)(f_3), \\ 4. & L_4(f) & = (f_1)(f_2 + 1)(f_3 + 1), \\ 5. & L_5(f) & = (f_1)(f_2 + 1)(f_3), \\ 6. & L_6(f) & = (f_1)(f_2)(f_3 + 1), \\ 7. & L_7(f) & = (f_1)(f_2)(f_3), \end{array} \tag{29}$$

и подставляя (28) в (29), получим разложение полиномов Лагранжа $L_k(f)$ по полиномам Лагранжа $L_j(x) = L_j$

$$\begin{array}{lll}
0. & L_0(f) = & (a_0+1)(b_0+1)(c_0+1)L_0 + (a_1+1)(b_1+1)(c_1+1)L_1 + \\
 & & +(a_2+1)(b_2+1)(c_2+1)L_2 + (a_3+1)(b_3+1)(c_3+1)L_3, \\
1. & L_1(f) = & (a_0+1)(b_0+1)(c_0)L_0 + (a_1+1)(b_1+1)(c_1)L_1 + \\
 & & +(a_2+1)(b_2+1)(c_2)L_2 + (a_3+1)(b_3+1)(c_3)L_3, \\
2. & L_2(f) = & (a_0+1)(b_0)(c_0+1)L_0 + (a_1+1)(b_1)(c_1+1)L_1 + \\
 & & +(a_2+1)(b_2)(c_2+1)L_2 + (a_3+1)(b_3)(c_3+1)L_3, \\
3. & L_3(f) = & (a_0+1)(b_0)(c_0)L_0 + (a_1+1)(b_1)(c_1)L_1 + \\
 & & +(a_2+1)(b_2)(c_2)L_2 + (a_3+1)(b_3)(c_3)L_3, \\
4. & L_4(f) = & (a_0)(b_0+1)(c_0+1)L_0 + (a_1)(b_1+1)(c_1+1)L_1 + \\
 & & +(a_2)(b_2+1)(c_2+1)L_2 + (a_3)(b_3+1)(c_3+1)L_3, \\
5. & L_5(f) = & (a_0)(b_0+1)(c_0)L_0 + (a_1)(b_1+1)(c_1)L_1 + \\
 & & +(a_2)(b_2+1)(c_2)L_2 + (a_3)(b_3+1)(c_3)L_3, \\
6. & L_6(f) = & (a_0)(b_0)(c_0+1)L_0 + (a_1)(b_1)(c_1+1)L_1 + \\
 & & +(a_2)(b_2)(c_2+1)L_2 + (a_3)(b_3)(c_3+1)L_3, \\
7. & L_7(f) = & (a_0)(b_0)(c_0)L_0 + (a_1)(b_1)(c_1)L_1 + \\
 & & +(a_2)(b_2)(c_2)L_2 + (a_3)(b_3)(c_3)L_3.
\end{array} \tag{30}$$

При получении разложения (30) учтено (5) и разложение единицы по L_j (10).

Зависимость общего вида между f_1, f_2, f_3 имеет вид:

$$F(f_1, f_2) = \sum_{k=0}^{7} A_k L_k(f) = 0. \tag{31}$$

Подставляя (30) в (31), получим разложение $F(f_1, f_2, f_3)$ по полиномам Лагранжа L_j

$$\begin{array}{ll}
F(f_1, f_2) = & \sum_{k=0}^{7}[(a_k+1)(b_k+1)(c_k+1)A_0 + \\
 & +(a_k+1)(b_k+1)(c_k)A_1 + (a_k+1)(b_k)(c_k+1)A_2 + \\
 & +(a_k+1)(b_k)(c_k)A_3 + (a_k)(b_k+1)(c_k+1)A_4 + \\
 & +a_k(b_k+1)c_kA_5 + a_kb_k(c_k+1)A_6 + a_kb_{kk}A_7] = 0.
\end{array} \tag{32}$$

В силу линейной независимости L_k из (32), будем иметь, что:

$$\begin{array}{l}
(a_k+1)(b_k+1)(c_k+1)A_0 + \\
+(a_k+1)(b_k+1)(c_k)A_1 + (a_k+1)(b_k)(c_k+1)A_2 + \\
+(a_k+1)(b_k)(c_k)A_3 + (a_k)(b_k+1)(c_k+1)A_4 + \\
+a_k(b_k+1)c_kA_5 + a_kb_k(c_k+1)A_6 + a_kb_{kk}A_7 = 0.
\end{array} \tag{33}$$

Система уравнений (35) – это система для 20 переменных из $GF(2)$: $a_0, \ldots, c_3$, $A_0, \ldots, A_7$, состоящая из 8 уравнений. Умножая (33) последовательно на $(a_k+1)(b_k+1)(c_k+1)$, $(a_k+1)(b_k+1)(c_k)$, $(a_k+1)(b_k)(c_k+1)$,

$(a_k+1)(b_k)(c_k)$, $(a_k)(b_k+1)(c_k+1)$, $a_k(b_k+1)c_k$, $a_kb_k(c_k+1)$, a_kb_{kk} и пользуясь тем, что

$$a_k(a_k+1)=0; b_k(b_k+1)=0; c_k(c_k+1)=0, \tag{34}$$

вместо (33) будем иметь следующие уравнения:

$$\begin{array}{cc} (a_k+1)(b_k+1)(c_k+1) & A_0=0,\\ (a_k+1)(b_k+1)c_k & A_1=0,\\ (a_k+1)b_k(c_k+1) & A_2=0,\\ (a_k+1)b_kc_k & A_3=0,\\ a_k(b_k+1)(c_k+1) & A_4=0,\\ a_k(b_k+1)c_k & A_5=0,\\ a_kb_k(c_k+1) & A_6=0,\\ a_kb_kc_k & A_7=0. \end{array} \tag{35}$$

Система уравнений (35) содержит 32 уравнения для 20 переменных. Она имеет слишком большое число разных решений, чтобы их все перечислить. Поэтому ограничимся алгоритмом их построения. Для разных $k=0,1,2,3$ уравнения из (35) связаны друг с другом только через $A_l, l=0,\ldots,7$. Поэтому имеет смысл рассматривать сначала по отдельности все случаи допустимых различных A_l. Всего их $2^8=256$. Но из этого числа надо исключить случай, когда все $A_l=0$, так как тогда зависимость между f_1, f_2, f_3 вырождается в тавтологию $0=0$ и случай, когда все $A_l=1$, так как в этом случае, складывая все 8 уравнений (35), получим, что $1=0$. Остается рассмотреть оставшиеся 254 случая. Они распадаются на следующие:

- 8 систем, когда 1 равно одно из A_l,
- 28 систем, когда 1 равно два из A_l,
- 56 систем, когда 1 равно три из A_l,
- 70 систем, когда 1 равно четыре из A_l,
- 56 систем, когда 1 равно пять из A_l,
- 28 систем, когда 1 равно шесть из A_l,
- 8 систем, когда 1 равно семь из A_l.

Рассмотрим в качестве иллюстрации построение решения одного из первого и одного из седьмого случаев. Пусть $A_0=1$; $A_1=A_2=\cdots=A_7=0$. Тогда (35) сводится к 4-м уравнениям:

$$(a_k+1)(b_k+1)(c_k+1)=0. \tag{36}$$

Придавая последовательно a_k, b_k, c_k значения, равные 0 и 1, получим, что для любого k существуют 7 решений, которые удобно представить в виде таблицы.

Таблица. Решения уравнения (36).

N	a_k	b_k	c_k
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

Поскольку для любого k можно взять любые решения для оставшихся других значений k, то существует $7^4 = 2401$ разных решений.

Поскольку первому случаю отвечают 8 разных систем, то в первом случае есть $8 \cdot 2401 = 19\,216$ разных зависимостей между f_1, f_2, f_3.

Случаи 2–6 рассматривается аналогично. Наконец, посмотрим седьмой случай, например, когда $A_0 = A_1 = \cdots = A_6 = 1; A_7 = 0$. Система уравнений (35) будет иметь вид:

$$\begin{aligned}
&(a_k+1)(b_k+1)(c_k+1) = 0,\\
&(a_k+1)(b_k+1)c_k = 0,\\
&(a_k+1)b_k(c_k+1) = 0,\\
&(a_k+1)b_kc_k = 0,\\
&a_k(b_k+1)(c_k+1) = 0,\\
&a_k(b_k+1)c_k = 0,\\
&a_kb_k(c_k+1) = 0.
\end{aligned} \tag{37}$$

Складывая все уравнения (37), получим:

$$a_kb_kc_k + 1 = 0. \tag{38}$$

Уравнение (38) имеет единственное решение $a_n = b_n = c_n = 1$. Значит в седьмом случае существует $1 \cdot 8 = 8$ разных зависимостей между f_1, f_2, f_3.

Список литературы

1. *Гришко М. Е.* Один из возможных способов разбиения файла на буферы при булевом сжатии файлов // Математика и ее приложения : журн. Иван. мат. о-ва. 2010. Вып. 1(7). С. 25–28.

2. *Толстопятов А. А.* О возможности использования булевых уравнений для сжатия файлов // Вестн. Иван. гос. ун-та. 2003. Вып 3. С. 82–84.

3. *Толстопятов А. А.* Построение кодирующего уравнения при булевом сжатии файлов // Математика и ее приложения : журн. Иван. мат. о-ва. 2010. Вып. 1(7). С. 69–83.

Поступила в редакцию 26.11.2012.