Математика и её приложения: ЖИМО. 2010. Вып. 1 (7). С. 69 – 82.

УДК 512.54

А. А. Толстопятов[1,2]

Построение кодирующего уравнения при булевом сжатии файлов

Ключевые слова: булево сжатие, булевы полиномы.

Получены оценки значений параметров кодирующего полинома. Предложен алгоритм построения кодирующего полинома в двух наиболее вероятных частных случаях выбора системы порождающих полиномов. Рассмотрена возможность построения категории кодирующих множеств булевых полиномов. Рассмотрена целесообразность использования этой категории для построения такого разбиения файла на буферы, кодирующее уравнение которого имеет решение.

Keywords: boolean compress, boolean polynoms.

We obtain estimates of the parameters for the coding polynomial. We suggest also the algorithm, which construct the encoding polynomial for two cases of choice the generating polynomials. We consider the possibility of constructing the category of coding sets of Boolean polynomials. We examine the feasibility of using this category for construction of the partition the file on buffers, which encoding equation has the solution.

Предложенный в [2] подход к сжатию файлов, основанный на использовании булевых уравнений, начинается с разбиения файла на буферы. При фиксированном разбиении файла для построения кода требуется определить следующие величины.

1. Коэффициенты кодирующего уравнения.

2. Систему порождающих булевых полиномов.

3. Подстановку полиномов из системы порождающих вместо булевых переменных, от которых зависит кодирующий полином.

Все эти величины можно рассматривать как булевы переменные, для которых ищется кодирующее уравнение [2]. Этих переменных слишком много. Поэтому возникает задача о независимом определении параметров указанных выше трех типов. В определенном смысле эта задача начала решаться в [3]. Однако, полученное там разделение указанных переменных было неполным, а алгоритм его получения достаточно сложен.

В [6] было показано, что наиболее вероятное число независимых булевых полиномов из кодирующих поля принадлежности будут совпадать с длиной кортежей, на которые разбит файл. Это значит, что в качестве системы кодирующих булевых полиномов $\varphi_p(x_i)$ ($i = 1, 2, \ldots, n$; $p = 1, 2, \ldots, n$) можно взять или сами булевы переменные x_i, или n из L (L — число буферов, на которые разбит файл) булевых полиномов $f_l(x_i)$, $l = 1, \ldots, L$, кодирующих поле принадлежности. Во втором случае оставшиеся $L-n$ из коэффициентов $f_l(x_i)$ должны выражаться через n взятых в

[1]Ивановский государственный университет; E-mail: khash2@mail.ru.

[2]Работа выполнена при финансовой поддержке РФФИ (проект 10-07-00350а).

© А. А. Толстопятов, 2010

качестве системы порождающих. Такие выражения можно получить только при определенном разбиении файла на буферы. Так как это разбиение можно варьировать, то возникает задача об алгоритме построения нужного в указанном выше смысле разбиения. Возможный, хотя не оптимальный алгоритм построения такого разбиения, был предложен в [1]. В настоящей работе рассматривается задача о построении кодирующего уравнения, если в качестве системы порождающих булевых полиномов взята одна из двух следующих:

1) x_i, $i = 1, \ldots, n$;

2) f_{l_i}, $i = 1, \ldots, n$.

Наконец, затронутый выше вопрос о построении такого разбиения файла, чтобы соответствующее ему кодирующее уравнение имело решение, начинает обсуждаться с вопроса об описании всех разбиений файла и преобразовании их друг в друга. Предложенные ранее в [4] и [5] подходы к решению задачи о разбиении файла, обладают тем недостатком, что в [4] требовалось, чтобы сжимался каждый буфер, что вовсе не является необходимым условием сжатия всего файла. А намеченные в [5] два подхода, связанные с квантовой информатикой и теорией распознавания образов не были доведены до конструктивных алгоритмов. Если начать строить такие алгоритмы, то мы все равно упремся в вопрос об описании всех разбиений файла и их преобразования друг в друга. Так как каждое разбиение файла после факторизации буферов по повторяющимся кортежам — это множество булевых полиномов $f_l(x_i)$, $l = 1, \ldots, L$; $i = 1, \ldots, n$ (L — число буферов, которые могут меняться при изменении разбиения, а n — длина кортежа, которая одна и та же для всех разбиений), то изменение разбиения файла — это отображение одной системы булевых полиномов в другую. Если такие множества булевых полиномов рассматривать как объекты, а их отображения друг на друга как морфизмы, то все возможные разбиения файла на буферы могут быть описаны как категория множеств булевых полиномов, на которой и надо строить, согласно [5], задачу о максимуме коэффициента сжатия.

1. Постановка задачи

Пусть файл длиной N_Φ битов разбит на кортежи длиной n. Нужно так объединить эти кортежи в L буферов, причем l-й буфер ($l = 1, 2, \ldots, L$) содержит m_l кортежей, из которых s_l — разные, чтобы выполнялись два условия:

1) коэффициент сжатия файла $k > 1$,

2) кодирующее уравнение имеет решения.

Коэффициент k определяется формулой

$$k = \frac{N_\Phi}{N_k}\,, \tag{1}$$

где N_K — длина кода, причем

$$N_\Phi = n \sum_{l=1}^{L} m_l \tag{2}$$

и

$$N_k = 2^I + 2^n \cdot P + PI \cdot \log_2 P + \log_2 \prod_{l=1}^{L} \frac{C_{m_l-1}^{s_l-1} m_l!}{\prod_{k=1}^{s_l} n_k^l!}\ , \tag{3}$$

где I — число булевых переменных в кодирующем полиноме $F(e_k)$ ($k = 1, 2, \ldots, I$), P — число порождающих булевых полиномов φ_p ($p = 1, 2, \ldots, P$), n_k^l — число повторов k-го кортежа в l-м буфере, $k = 1, 2, \ldots, s$, $l = 1, 2, \ldots, L$. Первое условие $k > 1$ будет возможным, если выполняется неравенство

$$\frac{N_\Phi}{N_k} > 1. \tag{4}$$

Параметрами разбиения файла служат числа n, m, L, I, P. Эти параметры, как было показано в [4] должны удовлетворять неравенствам:

$$\frac{(2^{I-n} + P)n}{n^2 - I(1 + 2^{-n} \log_2 P)n + \log_2 n} < L < \frac{(N_\Phi 2^{-n} - I \cdot 2^{-n} \log_2 P - P)n}{(1 + I \cdot 2^{-n} \log_2 P)n - \log_2 n}. \tag{5}$$

Если выполнено (5), то будет выполняться (4).

Второе условие может быть сформулировано следующим образом. Каждому из L буферов ставится в соответствие булев полином $f_l(x_i)$ [1], такой, чтобы уравнение

$$f_l(x_i) = 0 \tag{6}$$

имело s_l решений, т. е. кортежей длиной n, входящих в l-й буфер. Всему файлу в соответствие ставится кодирующее уравнение

$$F(\varphi_p^l) = f_l(x_i), \tag{7}$$

где φ_p ($p = 1, 2, \ldots, I$) — порождающие булевы полиномы от булевых переменных x_i, а $F(e_k)$ ($k = 1, 2, \ldots, I$) — кодирующий полином от I булевых переменных e_k. Для l-го буфера вместо e_k берется последовательность из I порождающих полиномов φ_p^l, причем не все φ_1^l, φ_2^l, $\ldots$, φ_I^l разные. Нужно так выбрать систему порождающих булевых полиномов $\varphi_p(x_i)$ ($p = 1, 2, \ldots, P$; $i = 1, 2, \ldots, n$) и подстановки $\varphi_k^l = e_k^l$, чтобы L уравнений (7) имели решения, т. е. чтобы кодирующий полином $F(e_k)$ имел одни и те же коэффициенты для всех L буферов. Вообще говоря, из кодирующего уравнения (7) должны одновременно определяться и коэффициенты кодирующего полинома $F(e_k)$ и коэффициенты системы порождающих булевых переменных $\varphi_p(x_i)$ и все L подстановок $\varphi_k^l = e_k^l$. Однако,

[1] x_i — булевы переменные ($i = 1, 2, \ldots, n$).

как показано в [6], можно отделить задачу об определении $\varphi_p(x_i)$ от задачи определения $F(e_k)$ и φ_k^l, если взять в качестве $\varphi_p(x_i)$, как отмечено выше, или сами булевы переменные x_i, или n полиномов f_{l_i} $(i = 1, 2, \ldots, n)$. Отметим, что хотя в [6] и было показано, что первый случай выбора $\varphi_p(x_i)$ наиболее вероятен, однако там не было исследовано, при каком разбиении файла на буферы, т. е. для какой системы полиномов $f_l(x_i)$, возможен выбор $\varphi_p(x_i)$ $(p = 1, 2, \ldots, P)$ с $P < n$. Во втором случае, если разбиение файла на кортежи содержит все из 2^n разных кортежей, то всегда можно разбить на буферы так, чтобы в качестве $\varphi_p(x_i)$ можно было взять P булевых полиномов f_{e_p} $(p = 1, 2, \ldots, P)$, но $P = 2^n$, что, конечно, хотя и упрощает построение $F(e_k)$, но увеличивает N_k. Очевидно, что предложенный в [1] алгоритм разбиения файла на буферы не является оптимальным. Будем предполагать, что можно выбрать n полиномов f_{e_p} $(p = 1, 2, \ldots, n)$, т. е. если $P > n$, то имеет смысл выбрать в качестве $\varphi_p(x_i)$ сами x_i, а не f_{e_p}. Если считать, что $P = n$ в обоих указанных выше частных случаях выбора $\varphi_p(x_i)$, то вместо (4), с учетом (1), (2), (3) будем иметь для первого случая:

$$k = \frac{n\sum\limits_{l=1}^{L} m_l}{2^I + n(2^n + I\log_2 n) + \log_2 \prod\limits_{l=1}^{L} \dfrac{C_{m_l-1}^{s_l-1} m_l!}{\prod\limits_{k=1}^{s} n_k^l!}}; \tag{8}$$

а вместо (5) будем иметь:

$$\frac{(2^{I-n}+n)n}{n^2 - I(1+2^{-n}\log_2 n)n + \log_2 n)} < L < \frac{(N_\Phi 2^{-n} - I\cdot 2^{-n}\log_2 n - n)n}{(1 + I\cdot 2^{-n}\log_2 n)n - \log_2 n}. \tag{9}$$

Таким образом, задача которая обсуждается в этой работе, заключается в том, чтобы найти коэффициенты кодирующего полинома $F(e_k)$ и подстановки $\varphi_k^l = e_k^l$, если в качестве $\varphi_p(x_i)$ взять x_i или $f_{e_i}(x_i)$ $(i = 1, 2, \ldots, n)$ так, чтобы удовлетворялось кодирующее уравенение (7).

2. Сжатие алгебры полиномов Лагранжа

Кодирующий полином $F(e_k)$ удобно разложить по базису из полиномов Лагранжа $L_k^e(e_i)$ $(k = 0, 1, \ldots, 2^I - 1;\ i = 1, 2, \ldots, I)$:

$$F(e_i) = \sum_{k=0}^{2^I-1} a_k L_k^e(e_i), \tag{10}$$

где

$$L_k^e(e_i) = \prod_{j=1}^{I} L_{k_j}^e(e_j), \tag{11}$$

k_j — коэффициенты в двоичном коде натурального числа k,

$$k = \sum_{j=1}^{I} k_j \cdot 2^{j-1}, \tag{12}$$

а полиномы Лагранжа $L^e_{k_j}(e_j)$ от одной булевой переменной e_j имеют вид:

$$L^e_0(e_j) = e_j + 1, \quad L^e_1(e_j) = e_j. \tag{13}$$

Полиномы Лагранжа L^e_k образуют базис главных идеалов, они удовлетворяют соотношению:

$$L^e_j \cdot L^k_j = \delta_{jk} L^e_j. \tag{14}$$

При подстановке вместо e_i для l-го буфера из набора I полиномов $\varphi^l_p,\ p = 1, 2, \ldots, I$, возможны два случая: $P < I$ или $P \geq I$. В первом случае среди φ^l_p обязательно будут одинаковые. Во втором случае среди φ^l_p могут быть несколько одинаковых, хотя, возможно, что все они разные. После подстановки $e^l_p = \varphi^l_p$ кодирующий полином становится полиномом от $\varphi_p,\ p = 1, 2, \ldots, P$, и вместо разложения (10) будем иметь:

$$F(\varphi_p) = \sum_{k=0}^{2^P-1} b_k L^{\varphi}_k(\varphi_p). \tag{15}$$

Отметим, что как в разложении (10), так и в разложении (11), коэффициенты a_k и b_k не зависят от номера буфера l, а длины будут одними и теми же для всех L буферов. Так как полиномы $f_l(x_i)$ являются функциями системы порождающих булевых полиномов $\varphi_p(x_i)$, то имеет место разложение:

$$f_l(x_i) = \sum_{k=0}^{2^P-1} c^l_k L^{\varphi}_k(\varphi_p). \tag{16}$$

Тогда из (15) и (16) следует, что кодирующее уравнение (7) может быть записано так:

$$\sum_{k=0}^{2^P-1} \left[b_k L^{\varphi}_k(\varphi_p) + c^l_k L^{\varphi}_k(\varphi_p) \right] = 0. \tag{17}$$

В свою очередь, т. к. $\varphi^l_{p_l}$ ($l = 1, 2, \ldots, L;\ p_l = 1, 2, \ldots, P$) — некоторые из полиномов φ_p, то их тоже можно разложить по полиномам Лагранжа $L^{\varphi}_k(\varphi_p)$:

$$L^{\varphi}_k(\varphi^l_{p_l}) = \sum_{j=0}^{2^P-1} c^l_{kj} L^{\varphi}_j(\varphi_p). \tag{18}$$

Из (17) и (18) будем иметь:

$$\sum_{k=0}^{2^P-1} \left[b_k \sum_{j=0}^{2^P-1} C^l_{kj} L^{\varphi}_j(\varphi_p) + c^l_k L^{\varphi}_k(\varphi_p) \right] = 0. \tag{19}$$

Записав $L_k^{\varphi}(\varphi_p)$ в виде

$$L_k^{\varphi}(\varphi_{p_l}^l) = \sum_{j=0}^{2^P-1} \delta_{kj} L_j^{\varphi}(\varphi_p), \tag{20}$$

получим из (19) и (20):

$$\sum_{k=0}^{2^P-1} \sum_{j=0}^{2^P-1} \left[b_k C_{kj}^l(\varphi_p) + c_k^l \delta_{kj} \right] L_k^{\varphi}(\varphi_p) = 0. \tag{21}$$

Так как разложение булева полинома, равного нулю, по полиномам Лагранжа имеет нулевые коэффициенты, то внешняя сумма по k в (21) исчезает, и вместо (21) будем иметь:

$$\sum_{j=0}^{2^P-1} \left[b_k C_{kj}^l(\varphi_p) + c_k^l \delta_{kj} \right] = 0. \tag{22}$$

В уравнениях (22) коэффициенты C_k^l известны из (16), если заданы булевы полиномы $f_l(x_i)$ и система порождающих полиномов $\varphi_p(x_i)$. Коэффициенты $C_{k_j}^l$ известны из (14), если заданы φ_p и подстановки $\varphi_{p_l}^l$, $l = 1, 2, \ldots, L$, $p_l = 1, 2, \ldots, P$. Поэтому система булевых уравнений (18) — это уравнения для определения коэффициентов b_k кодирующего полинома $F(\varphi_p)$, рассматриваемого как булев полином от порождающих полиномов φ_p. Система (22) — это система из $2^P \cdot L$ булевых уравнений, которая взятием от этих полиномов дизъюнкции по $k = 0, 1, \ldots, 2^P - 1$ и по $l = 1, 2, \ldots, L$ сводится к одному булеву уравнению:

$$\bigvee_{k=0}^{2^P-1} \bigvee_{l=1}^{L} \sum_{j=0}^{2^P-1} \left[b_k C_{kj}^l + c_k^l \delta_{kj} \right] = 0. \tag{23}$$

Поскольку кодирующий полином F должен зависеть от l_k булевых переменных, $k = 1, 2, \ldots, I$, а не от φ_p, $p = 1, 2, \ldots, P$, то необходимо ответить на вопрос, как связаны коэффициенты a_k, $k = 0, 1, \ldots, 2^I - 1$, в разложении (10) с коэффициентами b_k, $k = 0, \ldots, 2^P - 1$, в разложении (15), т. к. их можно найти, решая (23). Для ответа на этот вопрос нужно выяснить, как полиномы Лагранжа $L_k^e(e_i)$ ($i = 1, 2, \ldots, I$; $k = 0, 1, \ldots, 2^I - 1$) из разложения (10) связаны с полиномами Лагранжа $L_j^{\varphi}(\varphi_p)$, ($j = 0, 1, \ldots, 2^P - 1$; $p = 1, \ldots, P$). Поскольку эта связь возникает при подстановке вместо e_k каких-то из порождающих полиномов φ_p,

$$e_i^l = \varphi_{p_i}^l, \tag{24}$$

то именно эти подстановки и превращают $L_k^e(e_i)$ в $L_j^{\varphi}(\varphi_p)$. Выше отмечалось, что нужно рассматривать два случая, когда $P < I$ или $P \geq I$. Поскольку в первом случае обязательно, а во втором возможно среди $\varphi_{p_k}^l$

будут одинаковые, то можно ограничиться именно этим случаем, т. е. по сути дела только первым. В этом случае, каждая пара одинаковых $\varphi^l_{p_k}$ ровно половину полиномов Лагранжа обращает в ноль. Так как и $L^e_k(e_i)$, и $L^\varphi_k(\varphi_p)$ образуют алгебру, то превращение $L^e_k(e_i)$ в $L^\varphi_k(\varphi_p)$ может быть названо сжатием алгебры полиномов Лагранжа. Сложение и умножение в этих алгебрах задается следующим образом:

$$L^e_i + L^e_k = (\delta_{jk} + 1)(L^e_j + L^e_k)\,, \tag{25}$$

$$L^e_i L^e_k = \delta_{jk} L^e_j\,, \tag{26}$$

где δ_{jk} — единичная матрица с элементами из поля $GF(2)$. Аналогично:

$$L^\varphi_i + L^\varphi_k = (\delta_{jk} + 1)(L^\varphi_j + L^\varphi_k)\,, \tag{27}$$

$$L^\varphi_i L^\varphi_k = \delta_{jk} L^\varphi_j\,. \tag{28}$$

Так как согласно (11) полиномы $L^e_k(e_i)$ представлены в виде произведений полиномов Лагранжа от одной переменной, хотя в разных множителях эти переменные разные, а из (13) следует, что

$$L^e_{k_i} = e_i + k_j + 1\,, \tag{29}$$

то из (11) и (29) будем иметь

$$L^e_k(e_i) = \prod_{j=1}^{I}(e_j + k_j + 1)\,, \tag{30}$$

где k_j дается формулой (12). При подстановке (24) будем иметь из (30):

$$L^e_k(e_i) = L^e_k(\varphi^l_{p_i}) = \prod_{j=1}^{I}(\varphi_j + k_j + 1)\,. \tag{31}$$

Из предположения $I > P$ вытекает, что в (31) хотя бы два φ^l_j с разными j одинаковы. Тогда произведение из (31) с такими $\varphi^l_q = \varphi^l_r,\ q \neq r$, будет равно

$$(\varphi^l_q + k_q + 1)(\varphi^l_r + k_r + 1) = (\varphi^l_q + k_q + 1)(k_q + k_r + 1)\,. \tag{32}$$

Но правую часть (32) можно записать как

$$(\varphi^l_q + k_q + 1)(k_q + k_r + 1) = L^\varphi_{k_q}(\varphi^l_q)(k_q + k_r + 1)\,. \tag{33}$$

Тогда из (31) и (33) получим

$$L^e_k(e_i) = \prod_{q=1}^{I-1} L^\varphi_{k_q}(\varphi^l_q)(k_q + k_r + 1)\,. \tag{34}$$

Так как из $P < I$ следует, что $I = P + \sigma$, то должно существовать σ одинаковых $\varphi^l_{k_i}$. Поэтому повторяя приведенные выше рассуждения σ

раз для каждой пары одинаковых $\varphi^l_{q_i}$ будем иметь вместо (34) следующее выражение:

$$L^e_k(e^l_i) = \prod_{j=1}^{P} L^{\varphi}_{k_j}(\varphi^l_{p_j}) \cdot \prod_{i=1}^{I-P} (q_i + r_i + 1)\,. \tag{35}$$

Если заметить, что

$$\prod_{j=1}^{P} L^{\varphi}_{k_j}(\varphi^l_{p_j}) = L^e_k(e^l_i)\,, \tag{36}$$

то (35) можно переписать в виде:

$$L^e_k(e^l_i) = L^{\varphi}_k(\varphi^l_{p_l}) \cdot \prod_{i=1}^{I-P} (q_i + r_i + 1)\,. \tag{37}$$

Поэтому

$$\prod_{i=1}^{I-P} (q_i + r_i + 1) = \bigvee_{i=1}^{I-P} (q_i + r_i) + 1\,, \tag{38}$$

а дизъюнкция равна 1 только если все $q_i + p_i = 0$, $i = 1, 2, \ldots, I - P$, что выполняется только если $q_i = p_i$; (37) означает, что при подстановке (24) полином Лагранжа $L^e_k(e_i)$ либо превращается в полином Лагранжа $L^{\varphi}_k(\varphi_p)$, либо обращается в ноль.

Так как для разных буферов подстановки (24) разные, то сжатие алгебры полиномов Лагранжа будет тоже разным для разных буферов. Каждое из таких сжатий само является алгеброй, но они не являются подалгебрами алгебры полиномов Лагранжа от n булевых переменных x_i, $i = 1, \ldots, n$. Однако, все такие сжатия являются подалгебрами кольца булевых полиномов от x_i. Поэтому полученным ниже формулы следует рассматривать как записанные именно в этом кольце.

Если в (10) положим $e_i \to e^l_i$, то

$$F(e^l_i) = \sum_{k=0}^{2^I-1} a_k L^e_k(e^l_i)\,. \tag{39}$$

Аналогично, полагая в (15) $\varphi_p \to \varphi^l_{p_i}$, получим, что

$$F(\varphi^l_{p_l}) = \sum_{k=0}^{2^P-1} b_k L^{\varphi}_k(\varphi^l_{p_i})\,. \tag{40}$$

Подстановка (24) делает левые части (39) и (40) одинаковыми. Но тогда приравнивая правые части, подставляя в полученное равенство (37) и заменяя $q_i \to q^l_i\,,\ r_i \to r^l_i\,$, получим

$$\sum_{k=0}^{2^I-1} \left\{ a_k \prod_{i=1}^{I-P} (q^l_i + r^l_i + 1) + b_k \right\} L^{\varphi}_l(\varphi^l_{p_i}) = 0\,. \tag{41}$$

В (41) и в правой части (40) P было заменено на I, т. к. $P < I$. Поскольку левая часть (41) есть разложение нуля по полиномам Лагранжа $L_k^{\varphi}(\varphi_{p_i}^l)$, то коэффициенты такого разложение равны нулю, а значит вместо (41) будем иметь:

$$a_k \prod_{i=1}^{I-P} (q_i^l + r_i^l + 1) + b_k = 0. \tag{42}$$

Формулы (42) дают искомую связь a_k и b_k, а формулы (23) позволяют найти b_k по заданному файлу и его разбиению. Таким образом, получен способ построения кодирующего полинома, т. е. определения коэффициентов a_k.

3. Построение кодирующего уравнения в случае, когда порождающие полиномы есть булевы переменные

Как отмечалось выше, наиболее важным случаем числа порождающих P является случай $P = n$, когда в качестве φ_p можно выбрать сами булевы переменные $x_i,\ i = 1, \ldots, n$. В этом случае описанная в предыдущем разделе процедура построения кодирующего полинома упрощается. Вместо (10) будем иметь

$$F(x_i) = \sum_{k=0}^{2^n-1} a_k L_k(x_i), \tag{43}$$

т. к. даже если $I > P = n$, то сжатие алгебры полиномов Лагранжа $L_k(e_i)$ $(i = 1, \ldots, I;\ k = 0, \ldots, 2^I - 1)$, когда несколько групп e_i равны одним и тем же x_i, даст алгебру полиномов Лагранжа $L_k(x_i)$. Именно поэтому в (15) $b_k = a_k$, а в формуле (18) $c_{k_j}^l = \xi_k^l \delta_{k_j}$ и она приобретает вид:

$$L_k(x_{i_l}^l) = \xi_k^l L_k(x_i), \tag{44}$$

где $\xi_k^l = 1$ или 0 в зависимости от того, входят ли в множество $x_{i_l}^l$ все x_i или какие-то из них не входят. Тогда формула (43) превращается в тождество, а (23) принимает вид:

$$\bigvee_{k=0}^{2^n-1} \bigvee_{l=1}^{L} \sum_{j=0}^{2^n-1} (a_k \xi_k^l + c_k^l)\delta_{kj} = 0. \tag{45}$$

Параметры c_k^l меняются при изменении разбиения файла и дальнейшая задача заключается в том, чтобы получить ответ на следующий вопрос. Существуют ли для данного файла такие разбиения, чтобы уравнение (45), как уравнение для определения a_k, имело решение, или нет? В определенном смысле уравнение (45) и дает критерий существования такого разбиения.

4. Построение кодирующего уравнения в случае, когда порождающие полиномы являются полиномами, кодирующими поле принадлежности

Хотя вопрос о существовании разбиения файла, такого, что в качестве φ_p можно взять P полиномов $f_l(x_i),\ l = 1, \ldots, L$, остается открытым, это не мешает рассмотреть процедуру построения кодирующего уравнения в этом случае. Для упрощения формул будем считать, что порождающими $\varphi_p,\ p = 1, \ldots, P$, служат первые $f_l,\ l = 1, \ldots, L$. Это предположение не уменьшает общности рассмотрения, т. к. всегда можно ввести перестановку буферов, которая сведет общий случай к рассматриваемому.

Этот случай реализуется тогда, когда можно найти такое разбиение файла, что оказывается возможным решить P из L уравнений

$$f_l = f_l(x_i) \quad (i = 1, \ldots, n;\ l = 1, \ldots, L) \tag{46}$$

относительно x_i. Мы предполагаем, что это первые P из L уравнений (46). Таким образом, имеем:

$$x_i = x_i(f_p), \quad p = 1, 2, \ldots, P. \tag{47}$$

Тогда, подставляя (47) в (46), получим

$$f_l = f_l(x_i(f_p) = f_l(f_p)), \quad l = P+1, P+2, \ldots, L,\ p = 1, \ldots, P. \tag{48}$$

Если из f_p построить полиномы Лагранжа $L_l^f(f_p)$, $k = 0, 1, \ldots, 2^P - 1$, то известные зависимости (48), т. е. булевы полиномы $f_l(f_p)$, можно разложить по $L_k^f(f_p)$:

$$f_l(f_p)) = \sum_{k=0}^{2^P-1} C_k^l L_k^f(f_p). \tag{49}$$

В уравнении (10), чтобы выполнялось (7), принято вместо e_i^l, $(i = 1, \ldots, I;\ l = 1, \ldots, L)$ подставить какие-то из f_p, $i = 1, 2, \ldots, P$. Поэтому при $l = 1, 2, \ldots, P$ должны получаться f_p, а при $l = P+1, \ldots, L$ — зависимости (48), причем и те и другие получаются разложением по полиномам Лагранжа $L_k^f(f_p)$ $(k = 0, 1, \ldots, 2^P - 1;\ p = 1, 2, \ldots, P)$; можно утверждать, что $I > P$, а вместо (10) будем иметь:

$$F(f_p)) = \sum_{k=0}^{2^P-1} a_k L_e^f(e_i)\,. \tag{50}$$

Чтобы найти коэффициенты a_k нужно удовлетворить уравнениям (7), заменяя e_i на $f_{p_i}^l$, $l = 1, 2, \ldots, L$. При этом будем иметь:

$$f_l = \sum_{k=0}^{2^P-1} a_k L_k^e(f_{p_i}^l), \quad \text{если} \quad l = 1, 2, \ldots, P, \tag{51}$$

а также

$$f_l(x_i(f_p)) = \sum_{k=0}^{2^P-1} a_k L_k^e(f_{p_i}^l), \quad \text{если} \quad l = P+1, \ldots, L. \tag{52}$$

Поэтому при замене $e_i \to f_{p_i}^l$ вследствие $I > P$ какие-то из e_i обязательно окажутся одинаковыми, и обязательно произойдет сжатие алгебры полиномов Лагранжа $L_k^e(e_i)$ в алгебру $L_k^f(f_p)$.

Уравнения (51) не налагают никаких ограничений на a_k, поскольку полиномы f_l, $l = 1, 2, \ldots, P$, как порождающие, должны храниться в коде общего поля. Поэтому остается только $L-P$ уравнений (52). Раскладывая левые части (52) по полиномам Лагранжа $L_k^f(f_p)$, $p = 1, \ldots, P$, определим c_k^l:

$$f_l(x_i(f_p)) = \sum_{k=0}^{2^P-1} c_k^l L_k^e(f^p). \tag{53}$$

Поскольку, так же, как и в предыдущем разделе, система порождающих известна, а не определяется из кодирующего уравнения, то, видоизменяя (50) и (51), получим, что сжатие алгебры полиномов Лагранжа задается соотношениями

$$L_k^f(f_{p_i}^l) = \xi_k^l L_k^l(e_i), \tag{54}$$

а уравнение для определения коэффициентов кодирующего полинома будет иметь вид:

$$\bigvee_{k=0}^{2^n-1} \bigvee_{l=1}^{L} \sum_{j=0}^{2^n-1} (a_k \xi_k^l + c_k^l)\delta_{kj} = 0. \tag{55}$$

5. Задача о разбиении файла на буферы и категория множеств булевых полиномов

Все рассмотренные выше процедуры построения коэффициентов кодирующего уравнения предполагали, что эти коэффициенты строятся для фиксированного разбиения файла. Поскольку и при выборе в качестве порождающих x_i ($i = 1, \ldots, n$) или f_p ($p = 1, \ldots, P$) изменение разбиения файла меняло коэффициенты c_k^l в уравнениях (45) или (52), а параметры ξ_k^l, задающие сжатие алгебры полиномов Лагранжа, оставались свободными, то возникает вопрос об описании уравнения для a_k не для фиксированного разбиения файла, а для всего множества возможных разбиений. Не претендуя на полное решение этого вопроса, в этом разделе мы лишь наметим путь к такому решению. Пусть файл из N_Φ битов разбит на кортежи по n битов, а эти кортежи объединены в буферы, содержащие по m_l кортежей, $l = 1, 2, \ldots, L$. При одном и том же файле последовательность чисел m_l однозначно приводит к множеству булевых полиномов

$\{f_l\}, l = 1, \ldots, L$, которые получаются после вычеркивания из каждого буфера повторяющихся кортежей. Если рассматривать разбиения с разными числами буферов, то существуют $2^{[N_\Phi/n]}$ различных разбиений. Все они могут быть получены друг из друга с помощью одной операции передвижения границы вправо на один кортеж или передвижением границы между этими буферами влево. Если первая операция произведена m_{l+1} раз или вторая m_l раз, то границы между l-м буфером и $(l+1)$-м сливаются с границами между $(l+1)$-м и $(l+2)$-м буфером, и $(l+1)$-й буфер пропадет. Аналогично, повторение второй операции m_l раз приводит к исчезновению l-го буфера. Значит и в первом и во втором случаях происходит изменение L на $L-1$. Поскольку при таком порядке перехода от одного разбиения к другому число буферов может только уменьшиться, то, т. к.

$$1 \leq L \leq \left[\frac{N_\Phi}{n}\right], \tag{56}$$

то начинать надо со случая разбиения файла на $L = \left[\frac{N_\Phi}{n}\right]$ буферов, каждый из которых содержит по одному кортежу. Описанная выше операция, примененная к любой из $\left[\frac{N_\Phi}{n}\right] - 1$ границ между буферами, приведет к разбиению на $L-1$ буфер. Существенно, что вводимая операция на множестве всех разбиений кодируемого файла является частичной, т. к. не любые два разбиения можно перевести друг в друг с помощью одной такой операции. При этом всегда существуют последовательность этих операций, которая переводит любое разбиение файла в любое другое. Это означает, что разные разбиения одного и того же файла можно рассматривать как объекты категории таких разбиений, а последовательность описанных выше операций — как морфизмы в такой категории. Поэтому каждое разбиение одного и того же файла однозначно порождает множество булевых полиномов $\{f_l\}$, $l = 1, \ldots, L$, а $L = 1, 2, \ldots, \left[\frac{N_\Phi}{n}\right]$, то категория разбиения файла однозначно индуцирует категорию множеств булевых полиномов. Эти категории различаются одним важным моментом. Категорию разбиений кодируемого файла можно представить таблицей с $2^{\left[\frac{N_\Phi}{n}\right]}$ строками, каждая из которых содержит $\left[\frac{N_\Phi}{n}\right]$ чисел:

$$\begin{array}{lccccc} 1. & 1 & 1 & 1 & \ldots & 1 \\ 2. & 0 & 2 & 1 & \ldots & 1 \\ & & \ldots & & & \\ 2^{\left[\frac{N_\Phi}{n}\right]}. & 0 & 0 & 0 & \ldots & \left[\frac{N_\Phi}{n}\right] \end{array} \tag{57}$$

Число в строке — это длина буфера, т. е. число входящих в него кортежей. Число ноль соответствует отсутствующему буферу. Каждая строчка — это отдельное разбиение файла. Переход от одной строчки к соседней соответствует передвижению границы между буферами на один кортеж. Поскольку от любой n-й строки можно всегда перейти к любой l-й просто последовательно проходя все разделяющие их строки, то существуют

цепочки описанных выше элементарных операций, осуществляющие переход от любого разбиения файла к любому другому. Однако, такой переход может осуществляться даже если какие-то строчки удалены из таблицы (57). Именно поэтому целесообразно рассматривать отдельные строки в (57) как объекты категории, а описанные выше цепочки — как морфизмы.

В случае же категории множеств булевых полиномов $\{f_l\}$, $l = 1, \ldots, \left[\frac{N_\Phi}{n}\right]$ возникает то существенное отличие, что после разбиения файла на буферы из каждого буфера надо удалить повторяющиеся кортежи. То есть вместо последовательности m_l, $l = 1, \ldots, L$, в случае категории разбиений, появляется последовательность s_l, $l = 1, \ldots, L$. Однако, для этой последовательности универсального соотношения, аналогичного (2), не существует. Поэтому построить таблицу типа (57) нетрудно для конкретного файла, но невозможно для любого. Однако, категорию с объектами $\{f_l\}$ можно описать, превратив множество этих объектов в решетку, введя частичный порядок по вложению. Причем имеет смысл сначала построить решетку из отдельных булевых полиномов, частично упорядочив ее по включению множества решений разных булевых полиномов, а потом, объединяя эти полиномы в множества, с учетом их упорядочения, построить категорию множеств булевых полиномов. Такая категория является универсальной в том смысле, что каждому файлу в ней соответствует конкретный набор объектов, каждый из которых задает одно разбиение файла. Морфизмы в этой категории дают траектории, по которым можно переходить от одного разбиения к другому.

Нерешенной проблемой при таком подходе к описанию выбора разбиения файла является следующая. На этой категории нужно задать функционал, а именно отображение разбиения файла в множество чисел, каковыми служит коэффициент сжатия для данного разбиения файла. Этот функционал дается формулой (8). Однако, в этой формуле помимо m_l и s_l нужно задать еще числа повторов n_l^k, для включения которых в категорию разбиений, если это делать универсальным образом, описание этой категории приходится сильно усложнять. Именно нерешенность этой задачи позволяет дать пока только схематичное описание развитого подхода для решения проблемы разбиения файла на буферы.

Список литературы

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

2. *Толстопятов А. А.* О возможности использования булевых уравнений для сжатия файлов // Вестник ИвГУ. – 2003. – Вып. 3. – С. 82–84.

3. *Толстопятов А. А.* Алгоритм кодирования и декодирования поля принадлежности при булевом сжатии файлов // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2008. – Вып. 1 (5). – С. 53–76.

4. *Толстопятов А. А.* Алгоритм разбиения файла на буферы при булевом сжатии // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2008. – Вып. 1 (5). – С. 77–88.

5. *Толстопятов А.А.* Возможные подходы к разбиению файла на буферы при булевом сжатии // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2009. – Вып. 1 (6). – С. 129–138.

6. *Толстопятов А.А.* Построение системы порождающих полиномов при булевом сжатии файлов // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2010. – Вып. 1 (7). – С. 59–68.

Поступила в редакцию 25.12.2010