Математика и её приложения: ЖИМО. 2008. Выпуск 1 (5). С. 77 – 88.

УДК 512.54

А. А. Толстопятов[1]

Алгоритм разбиения файла на буферы при булевом сжатии

Ключевые слова: кортеж, буфер, коэффициент сжатия.

Исследована зависимость коэффициента сжатия от параметров разбиения файла на кортежи и объединения кортежей в буферы. Получены ограничения на число буферов. Построен алгоритм разбиения файла на буферы.

We research the connection between compression factor of the file and parametres of splitting it on trains and associations of trains into buffers. We receive restrictions on number of buffers. Also we construct the algorithm of splitting file on buffers.

Первой задачей, с которой приходится сталкиваться при булевом сжатии файлов, является задача об объединении кортежей, на которые разбит файл. Длина кортежей n вряд ли может отличаться от 8 или 16, ибо при ее увеличении слишком быстро растет объем вычислений, которые надо провести для построения кода. Особенностью подхода к сжатию информации с использованием булевых уравнений, предложенному в [1], является то, что до введения общего поля в [4] вне зависимости от числа кортежей, входящих в буферы, код 1-го поля — поля принадлежности, имеет одну и ту же длину, равную 2^n. Но и после введения общего поля он оказался равным для всех буферов, а именно, его длина есть $I\log_2 P$, где I — число переменных, от которых зависит кодирующий полином $F(e_i)$, $i = 1, \ldots, I$, а P — число порождающих булевых полиномов φ_p, $p = 1, 2, \ldots, P$. Это значит, что если L — число кортежей в l-ом буфере, то всего кортежей будет $\sum\limits_{l=1}^{L} m_l$, а число способов объединить кортежи в буферы — $2^{\sum\limits_{l=1}^{L} m_l - 1}$

Или, если обозначить длину файла через $N_Ф$ и учесть, что $N_Ф = n\sum\limits_{l=1}^{L} m_l$, то число таких способов равно $2^{N_Ф\, n-1}$, т. е. экспоненциально растет с ростом длины файла. Поэтому любое ограничение способов объединения кортежей в буферы является существенным для алгоритма булева сжатия. Этой задаче и посвящена настоящая работа.

1. Постановка задачи

Если обозначить через n длину кортежа, через L — число буферов, через m_l — число кортежей в l-м буфере, $l = 1, \ldots, L$, через I — число

[1]Ивановский государственный университет; E-mail: khash2@mail.ru. Работа выполнена при финансовой поддержке РФФИ (проект 07-07-00155).

булевых переменных e_i, $i = 1, \ldots, I$, от которых зависит кодирующий полином $F(e_i)$, [1], через P — число порождающих булевых полиномов φ_P, $p = 1, \ldots, P$, через s^l — число различных кортежей в l-ом буфере, через $D^{s_l}_{m_l-s_l}$ — длину таблицы стандартных форм для l-го буфера [2], через

$$\sigma^l = \max_{k=1}^{D^{s_l}_{m_l-s_l}} \sigma^l_k$$

— максимальную длину таблицы перестановок k-й стандартной формы l-го буфера, то коэффициент сжатия k будет равен

$$k = \frac{n \sum_{l=1}^{L} m_l}{2^I + 2^n \cdot P + LI \log_2 P + \sum_{l=1}^{L} \left(\log_2 D^{s_l}_{m_l-s_l} + \log_2 \sigma^l + \log_2 \frac{m_l!}{\prod_{k=1}^{s_l} n^l_k!} \right)}. \tag{1}$$

Условие сжатия файла

$$k > 1 \tag{2}$$

будет выполняться только при определенных ограничениях на параметры n, L, m_l, s^l, P, I, n^l_k ($l = 1, \ldots, L$, $k = 1, \ldots, s^l$). Частично вопрос о таких ограничениях исследован в [3, 4]; однако, в [3] он рассматривался не для всего файла, а для отдельного буфера, а в [4] исследовался вопрос об ограничениях на параметры кода общего поля — членов $2^I + 2^n \cdot P$, и параметров всех кодов поля принадлежности $LI \log_2 P$. Если же рассматривать такие ограничения для всего файла, то открывается возможность найти определенное распределение вероятностей для разных объединений кортежей в буферы, а также поставить задачу о нахождении максимального k. Это первая задача, рассматриваемая в работе.

Вторая задача связана с нахождением ограничений на L. Без решения этой задачи приходится перебирать все возможные способы объединения кортежей в буферы. Поскольку полный перебор за приемлемое время невозможен, а значит самый мощный фактор адаптации — выбор m_l (см. [3]) не используется.

Наконец, третья задача, обсуждаемая в работе — это построение алгоритма, позволяющего “шевелить” границы между двумя буферами так, чтобы условие (2) как для отдельного буфера, так и для файла в целом не нарушалось.

Результаты, обсуждаемые в настоящей работе, ни в коем случае не могут рассматриваться как решающие оптимальным образом задачу о разбиении файла при булевом сжатии. В первую очередь потому, что не рассматривается критерий существования решений кодирующего уравнения [1, 4] для данного разбиения файла. Речь идет только о том, какие ограничения можно извлечь из условия (2).

2. Зависимость коэффициента сжатия от параметров объединения кортежей в буферы

Пусть $n_1^l, n_2^l, \ldots, n_{s_l}^l$ — числа повторов l-го буфера. Обозначим через $p_1^l, p_2^l, \ldots, p_{a_l}^l$ числа одинаковых n_k^l, $k = 1, 2, \ldots, s^l$, при условии, что в разные группы с числом членов p^{l_k}, $k = 1, \ldots, a_l$, входят разные n_l^k. Тогда наряду с ограничением

$$\sum_{l=1}^{L} n_k^l = m_l \tag{3}$$

будем иметь условие

$$\sum_{k=1}^{a_l} p_k^l = s_l. \tag{4}$$

Используя введенные обозначения перепишем формулу (1) в виде:

$$k = \frac{n \sum\limits_{l=1}^{L} m_l}{2^I + 2^n \cdot P + LI \log_2 P + \log_2 \prod\limits_{l=1}^{L} \dfrac{D_{m_l - s_l}^{s_l} s_l! n_l!}{\prod\limits_{k=1}^{n_k^l} n_k^l! \prod\limits_{k=1}^{a_l} p_k^l!}}. \tag{5}$$

Поэтому длина таблицы перестановок для стандартной формы l-го буфера есть

$$\sigma_l = \frac{s^l!}{\prod\limits_{k=1}^{a_l} p_k^l!}. \tag{6}$$

Обозначим через k_l коэффициент сжатия l-го буфера. Тогда

$$k_l = \frac{n m_l}{\frac{\alpha}{L} + \beta_l}, \tag{7}$$

где

$$\alpha = 2^I + 2^n \cdot P + LI \log_2 P, \tag{8}$$

$$\beta_l = \log_2 \prod_{l=1}^{L} \frac{D_{m_l - s_l}^{s_l} s_l! n_l!}{\prod\limits_{k=1}^{s_l} n_k^l! \prod\limits_{k=1}^{a_l} p_k^l!}. \tag{9}$$

Коэффициент сжатия всего файла k из (5) может быть выражен следующим образом:

$$k = \frac{n \sum_{l=1}^{L} m_l}{\alpha + \sum_{l=1}^{L} \beta_l}. \tag{10}$$

Выясним, как связан k из (10) с k_l из (7). Умножив (7) на знаменатель дроби из правой части (7) и просуммировав по l от 1 до L, получим:

$$\frac{\alpha}{L} \sum_{l=1}^{L} k_l + \sum_{l=1}^{L} \beta_l k_l = n \sum_{l=1}^{L} m_l. \tag{11}$$

С другой стороны, записав (10) в виде

$$k \left(\alpha + \sum_{l=1}^{L} \beta_l \right) = n \sum_{l=1}^{L} m_l, \tag{12}$$

приравняв левые части (11) и (12), ввиду равенства их правых частей, и разрешив полученное уравнение относительно k, будем иметь:

$$k = \frac{\frac{\alpha}{L} \sum_{l=1}^{L} k_l + \sum_{l=1}^{L} \beta_l k_l}{\alpha + \sum_{l=1}^{L} \beta_l}. \tag{13}$$

Если ввести обозначение

$$\gamma_l = \frac{\beta_l}{\alpha}, \tag{14}$$

то нетрудно увидеть, что γ_l есть отношение длины кода полей кратности и порядка к суммарной длине кода общего поля и всех полей принадлежности, причем последние для всех буферов равны. Из (8), (9) и (14) получим следующее выражение для γ_l:

$$k = \frac{\log_2 D^{s_l}_{m_l - s_l} + \log_2 \frac{\sigma_l!}{\prod_{k=1}^{a_l} p^l_k} + \log_2 \frac{m_l!}{\prod_{k=1}^{s_l} n^l_k!}}{2^I + 2^n \cdot P + IL \log_2 P}. \tag{15}$$

Обозначая через $< k_L >$ среднеарифметическое коэффициентов сжатия всех L буферов

$$< k_L >= \frac{1}{L} \sum_{l=1}^{L} k_l, \tag{16}$$

перепишем (13) следующим образом:

$$k = \frac{< k_L > + \sum\limits_{l=1}^{L} \gamma_l k_l}{1 + \sum\limits_{l=1}^{L} \gamma_l}. \tag{17}$$

Выясним, как связаны k и $< k_L >$, если считать, что сжимаются все буферы, т. е. выполнено условие

$$k_l > 1,\ l = 1, 2, \ldots, L. \tag{18}$$

Переписав (17) в виде:

$$k - < k_L > = \sum_{l=1}^{L} \gamma_l (k - k_l) \tag{19}$$

и заметив, что, в силу условия сжатия файла (2), выполняется неравенство

$$\sum_{l=1}^{L} \gamma_l (k - k_l) + < k_L > > \ 1, \tag{20}$$

получим, что

$$\sum_{l=1}^{L} \gamma_l (k - k_l) > 1 - < k_L > . \tag{21}$$

Но вследствие (18) получим, что $< k_L > > 1$, а значит из (21) будем иметь неравенство

$$\sum_{l=1}^{L} \gamma_l (k - k_l) < 0. \tag{22}$$

Но тогда из (19) и (22) получим, что

$$k < < k_L > . \tag{23}$$

Таким образом, установлено, что *если сжимаются все буферы, то коэффициент сжатия всего файла будет меньше среднеарифметического коэффициентов сжатия всех буферов.*

Чтобы ответить на вопрос, может ли k быть больше единицы, если не выполнено (18), т. е. не все буферы сжимаются, запишем (17) в виде

$$k = \sum_{l=1}^{L} W_l k_l, \tag{24}$$

где

$$W_l = \frac{1 + L\gamma_l}{L\left(1 + \sum\limits_{l=1}^{L} \gamma_l\right)}. \tag{25}$$

Интересно отметить, что W_l из (25) удовлетворяет двум условиям

$$0 < W_l < 1, \tag{26}$$

$$\sum_{l=1}^{L} W_l = 1. \tag{27}$$

Неравенство (26) выполняется вследствие того, что все $\gamma_l > 0$, $L > 0$, а значит и $W_l > 0$; так как $\gamma_l < \sum\limits_{l=1}^{L} \gamma_l$ и $L > 1$, то $W_l < 1$. Условие (27) доказывается так:

$$\sum_{l=1}^{L} W_l = \frac{\sum\limits_{l=1}^{L}(1 + L\gamma_l)}{L\left(1 + \sum\limits_{l=1}^{L} \gamma_l\right)} = \frac{L + L\sum\limits_{l=1}^{L} \gamma_l}{L\left(1 + \sum\limits_{l=1}^{L} \gamma_l\right)} = 1. \tag{28}$$

Условия (26) и (27) позволяют рассматривать W_l как вероятность того, что коэффициент сжатия l-го буфера будет равным k_l. Тогда (24) означает, что коэффициент сжатия всего файла есть среднее значение коэффициентов сжатия отдельных буферов, вычисленное по распределению вероятностей W_l. Представление коэффициента сжатия файла (24) открывает возможность построения квантового алгоритма поиска оптимального разбиения файла, например, с использованием квантового алгоритма поиска в неупорядоченной базе данных Гровера [6].

Пусть теперь вместо (18) выполняются неравенства:

$$\begin{aligned} &k_l > 1, \quad \text{если } l = 1, 2, \ldots, P, \\ &k_l < 1, \quad \text{если } l = p+1, p+2, \ldots, L. \end{aligned} \tag{29}$$

Тогда (24) можно записать так:

$$k - \sum_{l=1}^{P} W_l k_l = \sum_{l=p+1}^{L} W_l k_l < L - p. \tag{30}$$

Неравенство (30) справедливо вследствие неравенств $W_l < 1$ $(l = 1, \ldots, L)$ и $k_l < 1$ $(l = p+1, p+2, \ldots, L)$, из которых следует $W_l k_l < 1$. Но сумма $L - p$ членов, каждый из которых меньше единицы, будет меньше числа этих членов.

Записав (30) в виде

$$\sum_{l=1}^{P} W_l k_l > k - L + p > 0, \tag{31}$$

что справедливо в силу $W_l > 0,\ k_l > 1,\ l = 1, 2, \ldots, P$, получим из (31) ограничение на число буферов L при разбиении файла:

$$L < k + p. \tag{32}$$

С другой стороны, из (30) будем иметь:

$$L > k + p - \sum_{l=1}^{P} W_l k_l > k + p - \sum_{l=1}^{p} k_l. \tag{33}$$

Последнее неравенство в (33) вытекает из того, что $W_l < 1$. Неравенство (33) даст ограничение на число буферов L снизу. Объединяя (32) и (33), будем иметь:

$$k + p - \sum_{l=1}^{p} k_l < L < k + p. \tag{34}$$

Неравенство (32) дает и оценку для числа буферов p, коэффициент сжатия которых должен быть больше единицы:

$$p > L - k > L - \{k\}. \tag{35}$$

Неравенство (34) можно упростить, заметив, что

$$\sum_{l=1}^{p} k_l > p \cdot \min_{l=1}^{p} k_l. \tag{36}$$

Тогда, с учетом (36), вместо (34) будем иметь

$$k - p \cdot \left(\min_{l=1}^{p} k_l - 1 \right) < L < k + p. \tag{37}$$

Хотя (37) и даст ограничения на L, но они носят неконструктивный характер, поскольку в них входят величины k, k_l, $l = 1, \ldots, P$, которые можно определить только после построения кода. Поэтому возникает задача о поиске более конструктивных неравенств типа (37).

3. Ограничения на число буферов L

Чтобы получить ограничения на число допустимых буферов L, подставим α из (8) в выражение для коэффициента сжатия (10) и получим, что

$$k = \frac{n\sum\limits_{l=1}^{L} m_l}{2^I + 2^n \cdot P + LI\log_2 P + \sum\limits_{l=1}^{L}\beta_l}. \tag{38}$$

Чтобы получить ограничение на L снизу, введем обозначения

$$< m_L >= \frac{1}{L}\sum_{l=1}^{L} m_l, \qquad < \beta_L >= \frac{1}{L}\sum_{l=1}^{L}\beta_l \tag{39}$$

для среднеарифметических значений числа кортежей во всех буферах $< m_L >$ и длин кодов всех полей кратности и порядка $< \beta_L >$. Тогда вместо (38) будем иметь:

$$k = \frac{n < m_L > L}{2^I + 2^n \cdot P + L(I\log_2 P + < \beta_L >)}. \tag{40}$$

Условие сжатия файла (2) и (4) дают:

$$L > \frac{2^I + 2^n \cdot P}{n < m_L > - I\log_2 P - < \beta_L >}. \tag{41}$$

Чтобы получить ограничения на L сверху, запишем (38) в виде

$$k = \frac{N_\Phi}{2^I + 2^n \cdot P + L(I\log_2 P + < \beta_L >)}, \tag{42}$$

где N_Φ — длина файла. Тогда условие сжатия файла (2) и (42) дают:

$$L < \frac{N_\Phi - 2^I - 2^n \cdot P}{I\log_2 P + < \beta_L >}. \tag{43}$$

Объединяя (41) и (43), будем иметь:

$$\frac{2^I + 2^n \cdot P}{n < m_L > - I\log_2 P - < \beta_L >} < L < \frac{N_\Phi - 2^I - 2^n \cdot P}{I\log_2 P + < \beta_L >}. \tag{44}$$

Так как левая часть неравенства (44) должна быть меньше правой, то, записав это неравенство, и решив его относительно $< \beta_L >$, получим ограничение на последнюю величину:

$$< \beta_L >< \frac{n < m_L > (N_\Phi - 2^I - P2^n) - N_\Phi I\log_2 P}{N_\Phi + 2^{I+1} + P2^{n+1}}. \tag{45}$$

Чтобы неравенства (44) стали конструктивными, остается оценить входящие в них величины $< m_L >$ и $< \beta_L >$ так, чтобы в эти оценки не входило L. Усилим неравенство (44) заменив в нем $< m_L >$ на $m_{\max}$, а $< \beta_L >$ — на $\beta_{\min}$. В результате получим:

$$\frac{2^I + 2^n \cdot P}{n m_{\max} - I \log_2 P - \beta_{\min}} < L < \frac{N_\Phi}{I \log_2 P + \beta_{\min}}. \tag{46}$$

В [5] были получены оценки для m_l. Для наших целей достаточно взять грубую оценку:

$$\frac{2^n}{n} < m_l < 2^n. \tag{47}$$

Тогда

$$m = m_{\max} = 2^n. \tag{48}$$

Если ввести обозначения

$$\begin{aligned}
D^s_{m-s} &= \max_{l=1,\ldots,L} D^{s_l}_{m_l - s_l}\,, \\
s_l &= \max_{l=1,\ldots,L} s_l\,, \\
C^{s-1}_{m-1} &= \max_{l=1,\ldots,L} C^{s_l-1}_{m_l-1}\,, \\
m &= \max_{l=1,\ldots,L} m_l = 2^n\,,
\end{aligned} \tag{49}$$

то, как показано в [5],

$$\beta_l \geq m \log_2 m > \log_2(D^s_{m-s} s! m!) > \log_2 C^{s-1}_{m-1}. \tag{50}$$

Поэтому в качестве грубой оценки для $\beta_{\min} < \beta_l$, $l = 1, 2, \ldots, L$, можно взять:

$$\beta_{\min} = m_{\min} \log_2 m_{\min} = 2^n - \frac{2^n}{n} \log_2 n, \tag{51}$$

причем в (51), согласно (47), взято $m_{\min} = \dfrac{2^n}{n}$. Из последнего равенства в (49), (51) и (46) получим, что

$$\frac{(2^{I-n} + P)n}{n^2 - (I + I \cdot 2^{-n} \log_2 P)n + \log_2 n} < L < \frac{(N_\Phi 2^{-n} - I \cdot 2^{-n} \log_2 P - P)n}{(1 + I \cdot 2^{-n} \log_2 P)n - \log_2 n}. \tag{52}$$

В неравенство (52) входят только длина файла N_Φ, длина кортежа n и параметры общего поля — число булевых переменных I, от которых зависит кодирующий полином $F(e_i)$, $i = 1, \ldots, I$, и число порождающих булевых полиномов P. Параметры N_Φ и n известны заранее, а параметры I и P подбираются начиная с минимальных допустимых путем проверки существования решений кодирующего уравнения, а в случае отсутствия таковых, последовательным увеличением их на единицу. Поэтому неравенства (52) являются конструктивными.

4. Алгоритм разбиения файла на буферы

Рассмотрим простейший алгоритм разбиения файла на кортежи длиной n и объединения этих кортежей в L буферов с числами кортежей, равными m_l, $l = 1, \ldots, L$. Идея этого алгоритма основывается на оценках, полученных в [3], для интервала числа кортежей в буфере, когда коэффициент сжатия каждого буфера $k_l > 1$. Хотя в [3] есть и более точные оценки, чем те, которые будут использованы дальше в этой работе, но с целью упрощения рассматриваемого алгоритма, ограничимся простейшим случаем. В [3] было показано, что k_l достигает максимума, если взять m равным

$$m_l = m = 0,69 \cdot 2^n. \tag{53}$$

Далее, в [3] было показано, что $k_l > 1$, если m лежит в интервале

$$m_1^l < m_l < m_2^l, \tag{54}$$

причем в (54):

$$m_1^l > 0,324 \cdot 2^n; \quad m_2^l < 1,906 \cdot 2^n. \tag{55}$$

Потребовав дополнительно к (55), чтобы выполнялось (54), получим неравенства, ограничивающие m_1^l и m_2^l как сверху, так и снизу:

$$0,324 \cdot 2^n < m_1^l < 0,69 \cdot 2^n, \tag{56}$$

$$0,69 \cdot 2^n < m_2^l < 1,906 \cdot 2^n. \tag{57}$$

Неравенства (56) и (57) показывают что, если сначала разбить файл на кортежи, а потом объединить кортежи в буферы равной длины $m_l = 0,69 \cdot 2^n$ согласно (53), то границы этих буферов можно передвигать так, чтобы выполнялись неравенства (56) и (57). Тогда k_l будет оставаться все время больше единицы и, как следствие этого, будут выполнены условия (2) сжатия всего файла. Границы между буферами приходится передвигать, если исходное объединение кортежей в буферы равной длины из (53) дает кодирующее уравнение, не имеющее решений, а после передвижения границ эти решения могут появиться.

Задача построения разбиения файла путем передвижения границ между буферами заключается в том, чтобы получить ограничения на эти передвижения, не нарушающие неравенств (56) и (57).

Будем задавать положение начала буфера с номером l числом предшестсвующих кортежей N_l. Тогда первоначальное разбиение файла характеризуется числами

$$N_0 = 0;\ N_k = \sum_{l=1}^{k} m_l = km = k \cdot 0,69 \cdot 2^n. \tag{58}$$

При этом длина l-го буфера, т. е. число кортежей, входящих в этот буфер, будет равно

$$m_l = N_l - N_{l-1} = m = 0,69 \cdot 2^n. \tag{59}$$

Пусть мы передвинули границы между буферами так, что теперь в $(l-1)$-й буфер входит $\tilde{m}_l$ кортежей. При этом число буферов L считаем неуменьшившимся. Тогда

$$\tilde{N}_0 = 0;\ \tilde{N}_k = \sum_{l=1}^{k} \tilde{m}_l,\ k = 1,\, 2,\, \ldots,\, L. \tag{60}$$

Длина k-го буфера будет равна:

$$\tilde{m}_l = \tilde{N}_l - \tilde{N}_{l-1} = \sum_{l=1}^{k} \tilde{m}_l - \sum_{l=1}^{k-1} \tilde{m}_l. \tag{61}$$

Так как $\tilde{m}_l$ должны удовлетворять неравенствам

$$0,324 \cdot 2^n < \tilde{m}_l < 1,906 \cdot 2^n, \tag{62}$$

то для $\tilde{N}_k$ из (61) и (62) получаем следующие неравенства:

$$\tilde{N}_k > \tilde{N}_{k-1} + 0,324 \cdot 2^n, \tag{63}$$

$$\tilde{N}_k < \tilde{N}_{k-1} + 1,906 \cdot 2^n. \tag{64}$$

Из (63) следует, что

$$\begin{aligned} \tilde{N}_k > \tilde{N}_{k-1} + 0,324 \cdot 2^n > \tilde{N}_{k-2} + 2 \cdot 0,324 \cdot 2^n > \cdots > \\ > \tilde{N}_0 + k \cdot 0,324 \cdot 2^n = k \cdot 0,324 \cdot 2^n, \end{aligned} \tag{65}$$

а из (64) —

$$\begin{aligned} \tilde{N}_k < \tilde{N}_{k-1} + 1,906 \cdot 2^n < \tilde{N}_{k-2} + 2 \cdot 1,906 \cdot 2^n < \cdots < \\ < \tilde{N}_0 + k \cdot 1,906 \cdot 2^n = k \cdot 1,906 \cdot 2^n. \end{aligned} \tag{66}$$

Объединяя (65) и (66), получим

$$k \cdot 0,324 \cdot 2^n < \tilde{N}_k < k \cdot 1,906 \cdot 2^n. \tag{67}$$

Выполнение неравенства (67) гарантирует, что после передвижения границ все $k_l > 1$, $l = 1, \ldots, L$. Поэтому простейший алгоритм объединения кортежей в буферы выглядит следующим образом.

1. Выбираем n ($n = 8,\, 16,\, \ldots$).

2. Выбираем $m_l = m = 0,69 \cdot 2^n$. Тогда число буферов равной длины m будет $L = \dfrac{N_\Phi}{0,69 \cdot 2^n}$.

3. Определяем $P_{\min}$ и $I_{\min}$ так, как это описано в [4].

4. Для n, $P_{\min}$, $I_{\min}$ и L из п. 2 проверяем удовлетворяется ли неравенство (52).

5. Если (52) удовлетворяется, то записываем кодирующее уравнение из [4].

6. Решаем кодирующее уравнение так, как это описано в [4].

7. Если кодирующее уравнение имеет решение, то сжатие возможно.

8. Если кодирующее уравнение не имеет решения, то последовательно увеличиваем на единицу $P_{\min}$ и $I_{\min}$ и повторяем пп. 4–7.

9. Если в п. 4 неравенство (52) не удовлетворяется, то разбиваем файл на несколько частей так, чтобы для каждой из них (52) выполнялось.

10. Для каждой из частей файла из п. 9 выполняем пп. 5–8.

Алгоритм, рассмотренный в этом разделе работы, является простейшим и допускает улучшение в следующих направлениях.

1. Вместо неравенств (56) и (57) взять более точные оценки из [3].

2. Заменить условие $k_l > 1$, $l = 1, 2, \ldots, L$, на более слабое, определяемое неравенствами (37).

3. Исследовать, как влияет на разбиение файла то обстоятельство, что при построении кода поля принадлежности, из кортежей входящих в каждый буфер нужно удалить повторяющиеся кортежи, сохранив только по одному разному кортежу. Однако, при передвижении границ между буферами эта процедура удаления повторяющихся кортежей будет давать разные буферы, что будет сказываться на выполнении неравенств (67).

Список использованной литературы

1. *Толстопятов А. А.* О возможности использования булевых уравнений для сжатия файлов // Вестник ИвГУ. – 2003. – Вып. 3. – С. 82–84.
2. *Толстопятов А. А.* Вычисление длины поля кратности при булевом сжатии файлов // Вестник ИвГУ. – 2004. – Вып. 3. – С. 71–76.
3. *Толстопятов А. А.* Факторы адаптации при булевом сжатии файлов // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2007. – Вып. 1(4). – С. 79–92.
4. *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.
5. *Толстопятов А. А.* Возможность кодирования поля кратности одним числом // Математика и ее приложения: Журн. Иванов. матем. об-ва. – 2008. – Вып. 1(5). – С. 43–52.
6. *Grover L.* A fast quantum mechanical algorithm for database search // Proceedings of the Annual ACM Symphosium on Theory of Computing, 28. – N.-Y., 1996. – P. 212–219.

Поступила в редакцию 29.12.2008.