Математика и ее приложения : ЖИМО. 2013. Вып. 1 (10). С. 37–42

УДК 512.543

А. А. Толстопятов[1]

Критерий поиска разбиения файла на буферы, удовлетворяющего кодирующему уравнению с помощью теории категорий

Ключевые слова: разбиение файла, порождающие полиномы, теория категорий.

В терминах описания всех разбиений файла на буферы на языке теории категорий сформулирован критерий разбиения файла, удовлетворяющего кодирующему уравнению.

Key words: boolean compress, boolean polynomials, Category theory.

In terms of the description of all partitions on the file buffers in the language of category theory formulated criteria to split file encoding satisfying equation.

Предложенный в [2] подход к сжатию файлов, оснванный на поиске и использовании, вообще говоря, нелинейных зависимостей между булевыми полиномами, кодирующими поля принадлежности буферов, на которые разбит файл, сталкивается с проблемой поиска нужного разбиения. Всех разбиений экспоненциально много и поэтому полный из перебор невозможен. Одним из подходов к поиску нужного разбиения может служить использование теории категорий [1]. Эта возможность обсуждена в [5]. Действительно, переход от одного разбиения к другому можно рассматривать как морфизм в категории, объектами которой служат сами разбиения. Преимуществом такого подхода является рассмотрение не только всех возможных разбиений, но и их преобразований друг в друга. Есть и еще одна проблема при построении разбиения файла на буферы. А именно, помимо построения разбиения, которое задается последовательностью чисел m_l кортежей, входящих в l-й буфер, где $l = 1, \ldots, L$, а L – число буферов, нужно построить систему порождающих булевых полиномов φ_p и кодирующее уравнение. Вообще говоря, эти задачи связаны друг с другом и их совместное решение оказывается очень сложным, в смысле

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

экспоненциально большого числа параметров, которые нужно определить. В [3] были рассмотрены два частных случая отделяющие эти задачи друг от друга. Эти случаи, когда в качестве порождающих берется часть полиномов, кодирующих поля принадлежностей и случай, когда порождающие – сами булевы переменные.

Для первого случая в [4] была рассмотрена задача о разбиении файла на буферы. Поэтому казалось бы естественно ограничиться рассмотрением либо первого либо второго частного случая. Однако, для рассматриваемой в этой работе задачи общий случай, когда не фиксируется заранее система порождающих, оказывается проще, так как не требует разложения по полиномам Лагранжа булевых переменных x_i или части булевых полиномов $f_l(x_i)$. Эти разложения носят достаточно специальных характер и требуют введения новых параметров. Для упрощения рассматриваемой задачи, так как наши цели ввести в кодирующее уравнение параметры описывающие переход от одного разбиения файла к другому, и выяснить, как эти параметры влияют на существование кодирующего полинома, ограничиваемся случаем, когда оба разбиения имеют одно и то же число буферов. Это ограничение непринципиально, так как отказ от него рассмотрен в [5].

1. Разбиение файла на буферы и их преобразования

Если n – длина кортежа, на которые разбит файл, то необходимо рассмотреть n булевых переменных $x_i, i = 1, \ldots, n$. Разбиение кодируемого файла на L буферов – это объединение кортежей в L буферов. Оно задается числами $m_l, l = 1, \ldots, L$ кортежей, входящих в l-й буфер. Если в l-м буфере удалить повторяющиеся кортежи, так что их останется $s_l \leq m_l$, то эти s_l кортежей можно рассматривать как решение булева уравнения

$$f_l(x_i) = 0. \tag{1}$$

Для каждого файла задание разбиения числами m_l однозначно приводит к полиномам f_l. Обратное, однако, неверно. Если полиномы f_l разложить по полиномам Лагранжа

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

то полином f_l однозначно определяет α_{lj}, также как и наоборот α_{lj} однозначно определяют f_l. Поэтому разбиение данного файла может быть задано матрицей размером $L \times 2^n$ α_{lj}. Другое разбиение файла с другими $\widetilde{m_l}$ однозначно дает другие полиномы $\widetilde{f}_l(x_i)$, которые могут быть заданы разложением по полиномам Лагранжа:

$$\widetilde{f}_l = \sum_{j=0}^{2^n-1} \widetilde{\alpha}_{lj} L_j(x)\,. \tag{3}$$

Это разбиение задается матрицей $\widetilde{\alpha}_{lj}$. При таком подходе переход от одного разложения к другому — это переход от α_{lj} к $\widetilde{\alpha}_{lj}$. Как показано в [5] при фиксированном L этот переход задается двумя невырожденными матрицами: O_{ls} размерности $L \times L$ и R_{jk} размерности $2^n \times 2^n$ согласно:

$$\widetilde{\alpha}_{lj} = O_{ls}\alpha_{sk}R_{kj}. \tag{4}$$

Пара (O, R) образует группу, действующую транзитивно на всех разбиениях данного файла на одно и тоже число буферов L.

2. Кодирующее уравнение и его преобразование

Для построения кодирующего уравнения нужно задать систему порождающих булевых полиномов $\varphi_p(x_i), p = 1, \ldots, P$ и кодирующий полином $F(e_k), k = 1, \ldots, I$, зависящий от I булевых переменных e_k. Если для l-го буфера заменить e_k на порождающие φ_p:

$$e_k^l = C_{kp}^l \varphi_p\,, \tag{5}$$

где L матриц C_{kp}^l в каждой строке содержит одну единицу и $I - 1$ нулей, то кодирующее уравнение есть:

$$F(e_k^l) = f_l. \tag{6}$$

Отметим, что выбор кодирующего полинома F и системы порождающих φ_p никак не зависит от разбиения файла. Однако, если коэффициенты F и φ_p рассматривать как булевы переменные, то в зависимости от правой части (6) система булевых уравнений (6) либо будет иметь решение, либо нет. Таким образом, от разбиения файла зависят именно правые части (6).

Чтобы ввести в уравнение (5) матрицы O и R, запишем разложение обеих частей (6) по полиномам Лагранжа. Вводя полиномы Лагранжа $L_k(e), k = 0, \dots 2^I - 1$ согласно:

$$L_k(e) = \prod_{s=1}^{k} L_{k_s}(e_s) = \prod_{s=1}^{k} (e_s + k_s + 1)\,, \tag{7}$$

где:

$$k = \sum_{s=1}^{I} k_s 2^{s-1}, \tag{8}$$

можно задать кодирующий полином $F(e)$ коэффициентами его разложения по $L_k(e)$:

$$F(e) = \sum_{k=0}^{2^I-1} A_k L_k(e) = \sum_{k=0}^{2^I-1} A_k \prod_{s=1}^{k} (e_s + k_s + 1)\,. \tag{9}$$

Если задать порождающие φ_p их разложением по полиномам Лагранжа:

$$\varphi_p = \sum_{j=0}^{2^n-1} b_{pj} L_j(x) \tag{10}$$

и воспользоваться тем, что

$$\sum_{j=0}^{2^n-1} L_j(x) = 1, \tag{11}$$

то используя (2), (4), (6), (10), (11) можно привести (9) к виду:

$$\sum_{j=0}^{2^n-1} \left\{ \sum_{k=0}^{2^I-1} A_k \prod_{s=1}^{k} (e^l_{kp} b_{jp} + k_s + 1) + \alpha_{lj} \right\} L_j(x) = 0\,. \tag{12}$$

В силу

$$L_i L_j = \delta_{ij} L_j\,, \tag{13}$$

полиномы Лагранжа линейно независимы, а значит из (12) следует:

$$\sum_{k=0}^{2^I-1} \left[A_k \prod_{s=1}^{k} (e^l_{kp} b_{jp} + k_s + 1) \right] + \alpha_{lj} = 0\,. \tag{14}$$

В случае булевых уравнений (14) неизвестными являются A_k, e^l_{kp}, b_{jp}, а заданными величинами – α_{lj}. При преобразовании одного разбиения с полиномами (2) в другое – с полиномами (3), вместо (14) будем иметь согласно (4):

$$\sum_{k=0}^{2^I-1}\left[A_k\prod_{s=1}^{k}(e^l_{kp}b_{jp}+k_s+1)\right]+O_{ls}\alpha_{sk}R_{kj}=0\,. \tag{15}$$

3. Критерий существования кодирующего уравнения и его решений

В системе (15), так как $l=1,\dots,L$, а $j=0,\dots,2^n-1$, 2^nL булевых уравнений. Но система булевых уравнений всегда может быть сведена к одному уравнению. Для этого достаточно от (15) взять дизъюнкцию всех уравнений системы. Тогда вместо (15) будем иметь:

$$\bigvee_{j=0}^{2^n-1}\bigvee_{l=1}^{L}\left\{\sum_{k=0}^{2^I-1}\left[A_k\prod_{s=1}^{k}(e^l_{kp}b_{jp}+k_s+1)\right]+O_{ls}\alpha_{sk}R_{kj}\right\}=0\,. \tag{16}$$

Так как

$$\bigvee_{k=1}^{n}f_k=\prod_{k=1}^{n}(f_k+1)+1, \tag{17}$$

то (16) можно придать вид:

$$\prod_{j=0}^{2^n-1}\prod_{l=1}^{L}\left\{\sum_{k=0}^{2^I-1}\left[A_k\prod_{s=1}^{k}(e^l_{kp}b_{jp}+k_s+1)\right]+O_{ls}\alpha_{sk}R_{kj}+1\right\}+1=0\,. \tag{18}$$

Уравнение (18) содержит в качестве дополнительных булевых переменных L^2 элементов O_{ls} и 2^{2n} элементов R_{kj}. Подбор этих параметров такой, чтобы левая часть (18) не вырождалась в единицу и дает искомый критерий.

Список литературы

1. *Букур И., Деляну А.* Введение в теорию категорий и функторов. М. : Мир, 1972. 190 с.

2. *Толстопятов А. А.* О возможности использования булевых уравнений для сжатия файлов // Вестник ИвГУ. 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. Иваново : ИвГУ, 2010. Вып. 1(7). С. 69–82.

4. *Толстопятов А. А.* Разбиение файла на буферы в случае, когда порождающие – часть множества булевых полиномов, кодирующих поля принадлежности // Математика и ее приложения: журн. Иван. мат. о-ва. Иваново : ИвГУ, 2011. Вып. 1(8). С. 105–112.

5. *Толстопятов А. А.* Применение теории категорий для разбиения файла на буферы // Математика и ее приложения: журн. Иван. мат. о-ва. Иваново : ИвГУ, 2011. Вып. 1(8). С. 121–128.

Поступила в редакцию 26.11.2012.