УДК 512.54

А. А. Толстопятов, С. И. Хашин[1]

Критерий разбиения файла на буферы, если система порождающих – сами булевы переменные

Ключевые слова: разбиение файла, порождающие булевы полиномы.

Если в качестве системы порождающих взять сами булевы переменные, то полиномы, кодирующие поля принадлежностией, должны быть симметрическими. Это значит, что кортежи, входящие во все буферы разбиения файла, должны получаться друг из друга перестановками входящих в них нулей и единиц.

Key words: boolean compress, boolean polynoms.

If the system generating take yourself Boolean variables, polynomials, encoding field accessories must be symmetrical. This means that the tuples that are members of all the buffers splitting the file should be obtained from each other permutations within them of ones and zeros.

Для построения разбиения файла на буферы [1] необходимо сформулировать критерии, которым должно удовлетворять это разбиение. Пока эта задача в общем виде еще очень далека от решения, имеет смысл получить такие критерии в частных случаях. Один из таких случаев – это случай, когда в качестве порождающих берутся сами булевы переменные.

Пусть $f_1, \dots, f_L$ – булевы полиномы от переменных $x_1, \dots, x_n$. Построить кодирующее уравнение (см. [2]) – это значит найти такой полином $F = F(e_1, \dots, e_P)$ от булевых переменных e_i и систему полиномов $\varphi_1, \dots, \varphi_P$ – порождающих, такие, чтобы при подстановке вместо e_p каких-то, возможно и одинаковых φ_p, выполнялись уравнения:

$$F(\varphi_{l_1}, \dots, \varphi_{l_P}) = f_l(x_1, \dots, x_n)\,. \qquad (1)$$

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

Рассмотрим отдельно три случая:

$$\begin{aligned} &1.\ P < n, \\ &2.\ P = n, \\ &3.\ P > n. \end{aligned} \tag{2}$$

Случай 3 можно сразу исключить, так как если $P > n$, то всегда в качестве φ_p можно взять $x_i, i = 1, \ldots, n$. Тогда $P = n$ и мы оказываемся уже в условиях случая 2. Именно он и будет рассмотрен в настоящей работе.

Случай 1, наиболее интересный с точки зрения приложений, требует дополнительных исследований и будет рассмотрен в отдельной статье.

Случай 2 интересен тем, что, во-первых, такие полиномы могут быть подобраны для любой системы исходных многочленов f_l, а во-вторых, он наиболее прост, так как система порождающих уже дана и ее не надо строить. При рассмотрении уравнений (1) нужно рассмотреть два разных случая:

$$\begin{aligned} &1.\ \text{Среди } x_{l_1}, \ldots, x_{l_n} \text{ нет одинаковых,} \\ &2.\ \text{Среди } x_{l_1}, \ldots, x_{l_n} \text{ есть одинаковые.} \end{aligned} \tag{3}$$

В первом случае (3), выбирая $P = n$ и $\varphi_p = x_p$, $p = 1, \ldots, n$, получаем, что разные уравнения в (1), т. е. уравнения при разном l должны получаться друг из друга перестановкой:

$$x_1, \ldots, x_n \to x_{l_1}, \ldots, x_{l_n}\,. \tag{4}$$

Это значит, что и полиномы f_l при разных l должны получаться друг из друга перестановкой булевых переменных x_i. Но в этом случае и решения булевых уравнений

$$f_l = 0, \tag{5}$$

дающие поля принадлежностей кортежей к l-му буферу, будут получаться

друг из друга заменой значений $x_i \in GF(2)$ согласно перестановке (4). Из этого следует критерий разбиения файла на буферы, если в качестве порождающих взяты булевы переменные x_i:

В разные буферы могут входить только такие кортежи, которые получаются из кортежей 1-го буфера с помощью перестановки x_i согласно (4).

Кодирующий полином $F(l_1, \dots, l_n) = F(x_1, \dots, x_n)$ можно взять в виде $F(x_1, \dots, x_n) = f_1(x_1, \dots, x_n)$.

Во втором случае из (3), когда среди x_i, $i = 1, \dots, n$ есть одинаковые, а значит, число различных x_i меньше n, полиномы $f_l(x_i)$ будут зависеть от меньшего числа x_i. Это значит, что при отождествлении $x_{l_2} = x_{l_1}$, переменная x_{l_2} в уравнении $f_l = 0$ не входит, а значит, она может принимать любое значение. Тогда критерий (6) модифицируется следующим образом:

В разные буферы могут входить только такие кортежи, которые получаются из кортежей 1-го буфера с помощью перестановки x_i согласно (4), или кортежи, в которых на фиксированном месте стоят как нули, так и единицы, а значения на всех остальных местах совпадают.

Очевидно, что в этом случае число кортежей в буфере должно быть четным.

Сформулированные критерии позволяют получать каждый последующий буфер из 1-го до тех пор, пока эти критерии будут выполняться. Если добавление следующего кортежа нарушает эти условия, то набор буфера прекращается и начинается набор следующего. Адаптация кода к конкретному файлу сводится в этом случае к выбору порождающих φ_p – это булевы переменные x_i, к выбору 1-го буфера и возможности обрывать набор буферов на любом месте до нарушения указанных выше условий.

Сформулированный алгоритм легко реализуется программно и поэтому может быть протестирован на реальных файлах. При этом с практической точки зрения более интересен случай, когда число порождающих меньше количества булевых переменных, от которых зависят булевы полиномы, кодирующие поля принадлежности. Однако в этом случае такого простого критерия не получается и поэтому он требует дополнительных исследований.

Список литературы

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

2. *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.