Математика и её приложения: ЖИМО. 2011. Вып. 1 (8). С. 121 – 128.

УДК 512.54

А. А. Толстопятов[1]

Применение теории категорий для разбиения файла на буферы

Ключевые слова: булево сжатие, булевы полиномы, теория категорий.

С использованием представления разбиения файла на буферы с помощью матриц показано, что все такие разбиения образуют категорию. Объектами этой категории служат разбиения с равным числом буферов, а морфизмами – отображения разбиения с одним числом буферов в разбиения с другим числом буферов. Построен функтор, отображающий эту категорию на себя. Показано, что все такие функторы образуют группу; получено описание этой группы.

Keywords: boolean compress, boolean polynoms, category theory.

Using the representaion of file splitting on buffers by matrixes, we show, that the set of all such partitions is some category. Objects of this category are sets of splittings with equal number of buffers; morphisms are mappings of these sets. We construct the funktor, mapping this category on itself; all such funktors form the group; we receive the description of this group.

Предложенный в [2] подход к сжатию файлов, основанный на использовании булевых уравнений, решения которых задают кортежи, входящие в тот или иной буфер, на которые разбит файл, обладает тем достоинством, что можно рассматривать разные разбиения файла на буферы. Поскольку алгоритм построения кода при булевом сжатии работает при любых разбиениях файла, то можно использовать любые из таких разбиений. Так как этих разбиений экспоненциально много, то булево сжатие имеет высокую адаптивность к сжимаемому файлу. Однако, это достоинство имеет и обратную сторону. А именно, вследствие экспоненциально большого числа разбиений задача поиска нужного разбиения или доказательство того, что для данного файла такого разбиения не существует, становится достаточно сложной. Сложность ее имеет две стороны. Во-первых, нужно для каждого разбиения файла выбирать систему порождающих булевых полиномов, записывать для этой системы кодирующее уравнение и проверять, имеет ли оно решение. Эта громоздкая процедура может быть упрощена, если в качестве системы порождающих выбрать часть булевых полиномов, кодирующих поля принадлежности. В [3] было показано, как в этом случае построить кодирующее уравнение. В [4] был предложен алгоритм описания разбиения файла при таком выборе порождающих с помощью матриц. Было рассмотрено преобразование этих матриц и показано, что в классе разбиений на одно и тоже число буферов эти преобразования образуют группу. Если же нужно описать переход от одного разбиения к другому, когда числа буферов отличаются друг от друга, то такие преобразования уже не будут иметь структуру группы, а должны рассматриваться как

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

морфизмы в категории [1]. Объектами этой категории служат разбиения конкретного файла с фиксированным числом буферов, и на этих объектах действуют группы, описанные в [4]. По видимому, именно такая категория и дает описание всех возможных разбиений файла и преобразований их друг в друга.

Существует точка зрения, — и ее придерживался, например, акад. А. И. Мальцев, что конструкция категории настолько обща и абстрактна, что на таком языке можно описать любую математическую структуру, и поэтому конструкция категории достаточно бесполезна. Поэтому первой задачей работы является исследование вопроса о роли категорий в математике. Второй задачей является построение категории разбиений файла, если эти разбиения описываются с помощью матриц, как это предложено в [4]. И, наконец, третьей задачей является переход описания всех разбиений файла и их преобразований друг в друга от языка теории категорий к языку теории групп. Для этого предлагается рассмотреть категорию как одну математическую структуру и ее преобразования на себя. Эти преобразования (функторы) образуют группу автоморфизмов данной категории.

1. Постановка задачи

Пусть файл разбит на N кортежей по n бит, и эти N кортежей объединены в L буферов. Пусть m_l $(l = 1, \ldots, L)$ — число кортежей в l-м буфере, s_l — число различных кортежей в l-м буфере, n_k^l $(k = 1, \ldots, s_l)$ — числа повторов в l-м буфере. Для конкретного файла числа n и m_l однозначно задают систему булевых полиномов $f_l(x_i),\ i = 1, \ldots, n$. Уравнения

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

определяет для каждого буфера те s_l кортежей, которые входят в этот буфер. Булево сжатие возможно тогда, когда существуют решения кодирующего уравнения (см. [3])

$$F(e_k^l) = f_l, \tag{2}$$

где $F(e_k^l)$ — кодирующий полином от I булевых переменных $l_k,\ k{-}1, \ldots, I$, которые для каждого l-го буфера выбираются из множества $\{\varphi_p\ \}$ порождающих булевых полиномов $(p = 1, \ldots, P)$. Коэффициент сжатия файла:

$$k = \frac{n\displaystyle\sum_{l=1}^{L} m_l}{2^I + 2^n P + LI\log_2 P + \log_2 \displaystyle\prod_{l=1}^{L} \frac{C_{m_l-1}^{s_l-1} m_l!}{\displaystyle\prod_{k=1}^{s} n_k^l!}}\,. \tag{3}$$

Разбиение файла должно удовлетворять двум условиям:

$$\begin{aligned} &1)\ k > 1,\\ &2)\ \text{уравнение (2) имеет решение.} \end{aligned} \tag{4}$$

Для того, чтобы найти разбиение, удовлетворяющее (4), нужно получить описание всех таких разбиений и их преобразования друг в друга. Задачу поиска нужного разбиения можно упростить, если в качестве φ_p выбрать часть $\{f_l\}$ [3]. Если записатьf_l в базисе Лагранжа

$$f_k = \sum_{j=0}^{2^n-1} \alpha_{kj} L_j(x_i), \tag{5}$$

то это разбиение задается матрицами α_{lj} (см. [4]). Изменение разбиения, как показано в [4], когда от (5) переходят к

$$\widetilde{f}_k = \sum_{j=0}^{2^n-1} \widetilde{\alpha}_{kj} L_j(x_i), \tag{6}$$

задается невырожденными квадратными матрицами O размерности $L \times L$ и R размерности $2^n \times 2^n$ согласно

$$\widetilde{\alpha} = O\alpha R. \tag{7}$$

Если есть три разбиения α, $\widetilde{\alpha}$, $\widetilde{\widetilde{\alpha}}$ и

$$\widetilde{\alpha} = O\alpha R\,, \quad \widetilde{\widetilde{\alpha}} = \widetilde{O}\widetilde{\alpha}\widetilde{R}\,, \quad \widetilde{\widetilde{\alpha}} = \widetilde{\widetilde{O}}\widetilde{\widetilde{\alpha}}\widetilde{\widetilde{R}}\,, \tag{8}$$

то

$$\widetilde{\widetilde{O}} = \widetilde{O}O, \quad \widetilde{\widetilde{R}} = \widetilde{R}R, \tag{9}$$

а значит, на парах (O, R) определено групповое умножение

$$(O, R) \cdot (\widetilde{O}, \widetilde{R}) = (\widetilde{\widetilde{O}}, \widetilde{\widetilde{R}}) = (\widetilde{O}O, \widetilde{R}R)\,. \tag{10}$$

Поэтому на множестве разбиений данного файла на фиксированное число буферов L действует группа, порожденная (O, R). Так как действие этой группы транзитивно, то все множество разбиений файла на L буферов (обозначм его через $\{L\}$) можно рассматривать как класс, задаваемый одним его представителем. Если $L_1 \neq L_2$, то отображения $\alpha_1 \in \{L\}$ в $\alpha_2 \in \{L\}$ может быть задано так:

$$\alpha_2 = O_{21}\alpha_1 R\,. \tag{11}$$

В (11), т. к. матрица α_2 имеет размерность $L_2 \times 2^n$, а α_1 — $L_1 \times 2^n$, матрица O_2 уже не квадратная, а имеет размерность $L_1 \times L_2$. Поэтому такие матрицы группы не образуют. Если все разбиения кодируемого файла на L_k буферов, где $L_k = k$ $(k = 1, \ldots, N)$ рассматривать как объекты $\mathrm{Ob}\,(L_k)$ в категории $\mathcal{K}(L)$ всех возможных разбиений файла на произвольное число буферов $L = 1, \ldots, L$, то отображения $\mathrm{Ob}\,(L_k) \to \mathrm{Ob}\,(L_s), k \neq s$, можно рассматривать как морфизмы в $\mathcal{K}(L)$. Эти морфизмы задаются парами $(O(k,l), R)$ где $O(k,l)$ — прямоугольная матрица размерности $l \times k$. Задачей этой работы является построение $\mathcal{K}(L)$ и группы функторов, отображающих $\mathcal{K}(L)$ на себя.

2. Почему категория?

Если предметом математики является математическая структура S,

$$S =< M \mid \rho_1,\, \ldots,\, \rho_s >, \tag{12}$$

где M — множество, а сигнатура $\rho_1,\, \ldots,\, \rho_s$ — отношения, задаваемые на декартовых степенях M,

$$\rho_i \subset M^i = M \times \cdots \times M, \tag{13}$$

то нетрудно убедиться в том, что вводя одно-единственное отношение ρ мы уже приходим к категории $\mathcal{K}(S)$ структур S. Усложнение сигнатуры принципиально ничего не меняет, поэтому рассмотрим простейший случай

$$S =< M \mid \rho >, \tag{14}$$

где $\rho \subset M^k$.

Множество M можно задавать не через отношение принадлежности, которое приводит к парадоксам, например, парадоксу Рассела, а через отображения множества на себя:

$$M \xrightarrow{g} M. \tag{15}$$

Если потребовать, чтобы g было биекцией, то все такие g образуют группу G. Группа G имеет инвариант — кардинальное число множества M. А так как M абстрактно, т. е. его элементы не конкретизированы, то два таких абстрактных множества могут различаться только кардинальными числами. Поэтому множество M само по себе определить нельзя, но можно определить пару (M, G). Снабдив M отношением ρ, превратим его в структуру (14). Отношение ρ выделит из группы G подгруппу G_ρ, которая будет сохранять ρ:

$$G_\rho \rho = \rho. \tag{16}$$

Для множества $G \backslash G_\rho$ отображений M на M вместо (16) будем иметь

$$g_1 \rho = \rho_1 \neq \rho, \tag{17}$$

где $g_1 \in G \backslash G_\rho$. Но тогда отображение g_1 порождает новую структуру S_{ρ_1}:

$$S_{\rho_1} =< M \mid \rho_1 > . \tag{18}$$

Вместе с этой структурой появляются и морфизмы

$$S_\rho \xrightarrow{g_1} S_{\rho_1}. \tag{19}$$

Условие (17) выделяет из G_ρ подгруппу G_{ρ_1}:

$$G_{\rho_1} \rho_1 = \rho_1, \tag{20}$$

действующую на S_{ρ_1}. Для множества $G \backslash G_{\rho_1}$ отображений M из M вместо (20) будем иметь

$$g_2 \rho = \rho_2 \neq \rho, \tag{21}$$

где $g_2 \in G \backslash G_{\rho_1}$. Но тогда отображение g_2 порождает новую структуру S_{ρ_2}:

$$S_{\rho_2} =< M \mid \rho_2 > \tag{22}$$

и множество морфизмов

$$S_\rho \xrightarrow{g_2} S_{\rho_2}. \tag{23}$$

Таким образом, изменение определения множества через принадлежность на определение через отображение M на M, — если на одном и том же M ввести отношение ρ, превратив его в структуру, — уже заменяет структуру S на категорию таких структур $\mathcal{K}(S)$. Поскольку разные отношения ρ, ρ_1, $\ldots$ вводят различие на M, то категории и появляются там, где такие различия есть.

Наконец, если $\mathcal{K}(S)$ понимать как пару $\{\mathrm{Ob}\,\mathcal{K}, \mathrm{Mor}\,\mathcal{K}\}$ множества объектов $\mathrm{Ob}\,\mathcal{K}$ и множества морфизмов $\mathrm{Mor}\,\mathcal{K}$, то можно определить функтор $F = (F_1,\, F_2)$ — биекцию на этой паре:

$$F_1 : \mathrm{Ob}\,\mathcal{K} \to \mathrm{Ob}\,\mathcal{K}, \tag{24}$$

$$F_2 : \mathrm{Mor}\,\mathcal{K} \to \mathrm{Mor}\,\mathcal{K}. \tag{25}$$

Биективность отображений (24) и (25) обеспечивает, что все они будут образовывать группу $G(F)$. Инвариантом этой группы будет категория $\mathcal{K}(S)$.

Эта общая схема превращения отдельной структуры S в категорию структур $\mathcal{K}(S)$ может быть применена к описанию различных разбиений файла. Действительно, если эти разбиения задавать с помощью матриц α, то разбиения файла с одинаковым числом буферов будут образовывать объекты в категории разбиений, на которых действует группа, инвариантом которой является число буферов, а отображения разбиений с разным числом буферов будут образовывать морфизмы в этой категории. С вышеизложенной точки зрения на категории всевозможные разбиения файла и их преобразования друг в друга образуют именно категорию.

3. Категория $\mathcal{K}(\alpha)$ матриц α_{ij}

Категория $\mathcal{K}$ есть тройка, включающая в себя следующее.

I. Класс объектов $\mathrm{Ob}\,\mathcal{K}$. Элементы этого класса будем обозначать $X,\, Y,\, Z,\, W,\, \ldots$.

II. Множество морфизмов $\mathrm{Mor}\,\mathcal{K}$. Элементы этого множества будем обозначать $\mathrm{Hom}\,(X,\, Y)$, а элементы множества $\mathrm{Hom}\,(X,\, Y)$ — через $\alpha,\ \beta,\ \ldots\ (\alpha :\ X \to Y,\, \beta :\ X \to Y,\, \ldots)$.

III. Отображение из $\mathrm{Hom}\,(X,\, Y) \times \mathrm{Hom}\,(Y,\, Z)$ в $\mathrm{Hom}\,(X,\, Z)$, называемое композицией. Если

$$\alpha : X \to Y, \quad \beta : Y \to Z, \quad \gamma : X \to Z,$$

то

$$\beta\alpha = \gamma; \tag{26}$$

при этом должны выполняться две аксиомы:

(а) если $\alpha : X \to Y$, $\beta : Y \to Z$, $\gamma : X \to Z$, то $\gamma(\beta\alpha) = (\gamma\beta)\alpha$;

(б) для $\forall\, X \in \operatorname{Ob}\mathcal{K}$ существует морфизм $\mathrm{id}_X : X \to X$ такой, что для $\forall\, Y \in \operatorname{Ob}\mathcal{K}$ и $\forall\, \alpha \in \operatorname{Hom}(X, Y)$ выполняются равенства

$$\alpha\, \mathrm{id}_X = \alpha \text{ и } \mathrm{id}_Y \alpha = \alpha. \tag{27}$$

Построить категорию — это значит описать объекты, морфизмы, композицию (26) и проверить аксиомы (а) и (б).

Обозначим через $\mathcal{K}(\alpha)$ категорию, описывающую все разбиения файла на буферы, если он содержит N котрежей по n бит. Объектами этой категории служат множества матриц α_{lj} $(l = 1, 2, \ldots, L;\ j = 0, \ldots, 2^n - 1)$. Так как L может меняться от 1 до N, то можно записать:

$$\operatorname{Ob}\mathcal{K}(\alpha) = \bigcup_{L=1}^{N} \bigcup_{l=1}^{L} \{\alpha_{lj}\}. \tag{28}$$

Если обозначить

$$X_L = \bigcup_{l=1}^{L} \{\alpha_{lj}\}, \tag{29}$$

то вместо (28) будем иметь:

$$\operatorname{Ob}\mathcal{K}(\alpha) = \bigcup_{L=1}^{N} X_L. \tag{30}$$

Морфизмы в $\mathcal{K}(\alpha)$ будут двух типов:

$$\begin{aligned} &\alpha_L : X_L \to X_L, \\ &\alpha_{L_1L_2} : X_{L_1} \to X_{L_2} \quad (L_1 \neq L_2). \end{aligned} \tag{31}$$

Вводя обозначения $\{\alpha_L\} = \operatorname{Hom}(X_L, X_L)$ и $\{\alpha_{L_1L_2}\} = \operatorname{Hom}(X_{L_1}, X_{L_2})$, можно заметить, что

$$\operatorname{Mor}\mathcal{K}(\alpha) = \bigcup_{L=1}^{N} \operatorname{Hom}(X_L, X_L) = \bigcup_{L_1=1}^{N} \bigcup_{L_2=1}^{N} \operatorname{Hom}(X_{L_1}, X_{L_2}). \tag{32}$$

$\operatorname{Hom}(X_L, X_L)$ задается парой матриц (O, R), согласно (7), причем матрица O размера $L \times L$, а матрица R размера $2^n \times 2^n$; $\operatorname{Hom}(X_{L_1}, X_{L_2})$ задается парой матриц (O, R) согласно (7), причем матрица O размера $L_1 \times L_2$, а R размера $2^n \times 2^n$.

Композиция двух морфизмов задается по разному для $\operatorname{Hom}(X_L, X_L)$ и $\operatorname{Hom}(X_{L_1}, X_{L_2})$, $\operatorname{Hom}(X_{L_2}, X_{L_3})$ и $\operatorname{Hom}(X_{L_1}, X_{L_3})$. В случае если $\alpha, \beta, \gamma \in \operatorname{Hom}(X_L, X_L)$, то

$$\alpha = (O_\alpha, R_\alpha), \quad \beta = (O_\beta, R_\beta), \quad \gamma = (O_\gamma, R_\gamma) \tag{33}$$

и композиция (26) задается согласно (10):

$$(O_\gamma, R_\gamma) = (O_\beta O_\alpha, R_\alpha R_\beta). \tag{34}$$

Композиция (34) удовлетворяет всем аксиомам группы и поэтому $\mathrm{Hom}\,(X_L,\, X_L)$ образует группу.

Если $\alpha \in \mathrm{Hom}\,(X_{L_1},\, X_{L_2})$, $\beta \in \mathrm{Hom}\,(X_{L_2},\, X_{L_3})$, $\gamma \in \mathrm{Hom}\,(X_{L_1},\, X_{L_3})$, то

$$\alpha = (O_{L_1L_2}, R_\alpha), \quad \beta = (O_{L_2L_3}, R_\beta), \quad \gamma = (O_{L_1L_3}, R_\gamma), \tag{35}$$

и композиция (26) задается согласно (10):

$$(O_{L_1L_3}, R_\gamma) = (O_{L_1L_2} O_{L_2L_3}, R_\beta R_\alpha). \tag{36}$$

Справедливость аксиом (а) и (б) следует из того, что композиции морфизмов заданы с помощью умножения матриц, а это умножение ассоциативно. В случае морфизмов $\mathrm{Hom}\,(X_L,\, X_L)$ матрицы $O_\alpha,\ O_\beta,\ O_\gamma,\ R_\alpha,\ R_\beta,\ R_\gamma$ являются квадратными и невырожденными, а значит

$$\mathrm{id}_X = (I_O, I_R), \tag{37}$$

где I_O — единичная матрица размерности $L \times L$, а I_R — единичная матрица размерности $2^n \times 2^n$. Таким образом, построенная $\mathcal{K}(\alpha)$ действительно является категорией.

4. Функторная группа

Произвольный функтор T отображает категорию $\mathcal{K}_1$ в $\mathcal{K}_2$, где, вообще говоря, $\mathcal{K}_1$ не совпадает с $\mathcal{K}_2$. Но нас будет интересовать только случай, когда

$$\mathcal{K} \to \mathcal{K}. \tag{38}$$

Функтор T есть пара (T_1, T_2), включающая в себя:

(I) отображение $T_1 : \mathrm{Ob}\,\mathcal{K}(\alpha) \to \mathrm{Ob}\,\mathcal{K}(\alpha)$,

(II) отображение $T_2 : \mathrm{Mor}\,\mathcal{K}(\alpha) \to \mathrm{Mor}\,\mathcal{K}(\alpha)$, сохраняющее композиции и тождественные морфизмы, такие, что

$$T_2(\beta\alpha) = T_2(\beta) \circ T_2(\alpha), \tag{39}$$

для любых морфизмов $\alpha : X \to Y$, $\beta : Y \to Z$ и

$$T_2(\,\mathrm{id}_X) = \,\mathrm{id}_{T_1X}, \tag{40}$$

для любого $X \in \mathrm{Ob}\,\mathcal{K}(\alpha)$.

Композиция T_1 в паре определяется тем, что

$$T_1(\alpha_{L_1}) = \alpha_{L_1}. \tag{41}$$

Поэтому T_1 можно задать матрицей порядка $L_1 \times L_1$, согласно (7).

Композиция T_2 в паре определяется тем, что

$$T_2(O_{L_1L_2}, R) = (\widetilde{O}_{L_1L_2}, \widetilde{R}), \tag{42}$$

поскольку $T_2 \quad \to \quad \mathrm{Hom}\,(X_{L_1}, X_{L_2}) \quad \to \quad \mathrm{Hom}\,(X_{L_1}, X_{L_2})$, а $(O_{L_1L_2}, R) \in \mathrm{Hom}\,(X_{L_1}, X_{L_2})$ и $(\widetilde{O}_{L_1L_2}, \widetilde{R}) \in \mathrm{Hom}\,(X_{L_1}, X_{L_2})$. Но тогда ее можно задать с помощью пары матриц

$$T_2 = (\widetilde{O}_{L_1L_2} O^{-1}_{L_1L_2}, R^{-1}\widetilde{R}), \tag{43}$$

т. к. из (43) и (36) вытекает (42). Условия (39) и (40) для пары (T_1, T_2), определяемой (7) и (43), проверяются непосредственно.

Можно определить композицию функторов

$$\widetilde{\widetilde{T}} = T\widetilde{T}, \tag{44}$$

где $T = (T_1, T_2)$, $\widetilde{T} = (\widetilde{T}_1, \widetilde{T}_2)$, $\widetilde{\widetilde{T}} = (\widetilde{\widetilde{T}}_1, \widetilde{\widetilde{T}}_2)$, а компоненты $T, \widetilde{T}, \widetilde{\widetilde{T}}$ определены, согласно (7) и (43). Поскольку компонента T_2 задается парой квадратных и невырожденных матриц, то умножение

$$\widetilde{\widetilde{T}}_2 = T_2\widetilde{T}_2, \tag{45}$$

можно задать умножением на этих парах матриц покомпонентно. Что касается композиции T_1, то в композиции

$$\widetilde{\widetilde{T}}_1 = T_1\widetilde{T}_1, \tag{46}$$

т. к. каждый из сомножителей в (46) задается парой квадратных и невырожденных матриц (O, R) из (7), умножение которых удовлетворяет аксиомам группового умножения. Нетрудно проверить, что умножение функторов (44) удовлетворяет аксиомам группы, и поэтому все такие функторы образуют группу.

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. Иванов. матем. об-ва. – 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. Иванов. матем. об-ва. – 2011. – Вып. 1 (8). – С. 105–112.

Поступила в редакцию 15.12.2011