

ПОИСК ЧАСТИЧНО ИЗВЕСТНОГО СЛОВА В СЛОВАРЕ

М. М. Трофимова

Исследуется проблема поиска частично известных слов в машинной памяти. Рассматривается важный частный случай, когда каждое слово состоит из двух частей, называемых первым и вторым ключами соответственно. Требуется находить слово, зная лишь один из ключей. Предполагается, что в словаре использовано A разных первых ключей, на каждый из которых приходится по r вторых ключей. Если в каждой ячейке машинной памяти хранится не более одного слова, т. е. нет склеиваний, то по заданному первому ключу можно найти слово за r шагов, а по второму — за A шагов. Предлагается простой алгоритм, позволяющий разместить без склеиваний любой словарь с двухключевыми словами, используя объем машинной памяти $A^{3/2}r$ — по порядку равный минимальному. Известные методы поиска Р. Райвеста, а также Д. Слепьяна и Д. Вулфа при размещении допускают склеивания и используют объем памяти Ar . Однако время поиска слова при таком размещении по первому или второму ключу не превосходит Ar , а не r или A .

Введение

Рассматривается процесс размещения информации в памяти компьютера с последующим возможно более быстрым извлечением ее. Часто перед пользователем стоит простая поисковая задача, когда в памяти хранится, например, словарь и нужно найти запись, соответствующую данному слову из этого словаря. Д. Кнут [1] описал различные алгоритмы для решения простой поисковой задачи. Однако важно также уметь решать более сложную задачу — находить запись, зная слово частично. Д. Кнут отмечает, что существующие методы поиска по частично известному слову неизмеримо хуже, чем методы для простой поисковой задачи.

Рассмотрим один из важных частных случаев задачи поиска записи в словаре по частично известному слову, когда предполагается, что слово состоит из двух подслов, называемых *первым* и *вторым* ключами. Необходимо разместить словарь в машинной памяти, т. е. каждому

слову требуется поставить в соответствие адрес ячейки машинной памяти, где будет храниться соответствующая ему запись. Адрес ячейки состоит из двух частей. Первая часть является функцией от первого ключа, а вторая — функцией от второго ключа. Размеры областей значений этих функций определяют объем машинной памяти, необходимой для размещения словаря.

Например, пусть имеется список студентов, состоящий из фамилий и имен. Предположим, что в списке нужно найти всех студентов с данной фамилией. Для этого нужно вычислить первую часть адреса и просмотреть в машинной памяти все соответствующие ячейки. Если размещение списка осуществлялось взаимно-однозначным на нем отображением, то записи ячеек, содержащие требуемую фамилию, и есть результат поиска. Если же при размещении встречались склеивания, т. е. в одну и ту же ячейку попадали несколько разных элементов списка, то необходимо дополнительно проводить перебор внутри каждой просматриваемой ячейки памяти.

Будем считать, что время поиска складывается из числа обращений в непустые ячейки и числа просмотренных внутри ячейки записей, тогда склеивания увеличивают время поиска. В настоящей статье изучается возможность размещения двухключевого словаря без склеиваний и с использованием объема памяти, близкого к объему словаря.

Дадим формальное описание задачи. Пусть даны два конечных множества X и Y , а также подмножество S их декартова произведения. Отображение на S осуществляется двумя функциями, заданными на X и Y соответственно. Какими должны быть размеры областей значений этих функций, чтобы можно было найти пару, дающую инъективное на S отображение? Очевидно, что достаточно иметь область объема $|X|$ для первой функции и область объема $\min\{|S|, |Y|\}$ для второй, но является ли это условие необходимым?

Д. Кнут [1] описал алгоритм, предложенный Р. Райвестом, в котором используется одна функция, определенная на объединении X и Y , с размером области значений, равным $\sqrt{|S|}$. Эта функция задает на S отображение в множество объема $|S|$. Это отображение допускает склеивания. Поэтому в большинстве списков время поиска по фамилии равно $\max\{\sqrt{|S|}, |Y|\}$, а по имени — $|X||Y|/\sqrt{|S|}$. Однако в наихудшем случае поиск сводится к исчерпывающему перебору всех записей, т. е. время поиска как по имени, так и по фамилии равно $|S|$.

Д. Слепяном и Д. Вулфом [4] рассматривалась задача кодирования двух зависимых источников, которую можно интерпретировать как размещение списка, в котором каждая фамилия встречается с одинаковым числом имен, допускающее склеивание небольшого числа элементов.

Полученный ими результат показывает, что доля склеивающихся элементов списка с ростом $|S|$ будет стремиться к нулю тогда и только тогда, когда размеры областей значений равны $|X|^{1+\varepsilon}$ и $(|S|/|X|)^{1+\varepsilon}$ соответственно, где $\varepsilon > 0$ может быть сколь угодно мало. В этом случае для большинства элементов время поиска по фамилии равно $|S|/|X|$, а по имени — $|X|$, но для $o(|S|)$ элементов время поиска как по имени, так и по фамилии в худшем случае увеличивается до $|S|$.

В настоящей работе предлагается простой алгоритм, с помощью которого для любого двухключевого словаря S можно найти инъективное отображение двумя функциями с областями значений не больше чем $|X|$ и $|Y|$ соответственно. Если в словаре на каждый из $|X|$ первых ключей приходится поровну вторых ключей, то алгоритм обеспечивает размещение без склеиваний, используя минимальный с точностью до порядка объем машинной памяти, равный $|S|\sqrt{|X|}$. Следует заметить, что реальное сокращение объема памяти от $|X||Y|$ до $|S|\sqrt{|X|}$ получается для словарей, в которых количество вторых ключей, соответствующих одному первому ключу, не превосходит $|Y|/\sqrt{|X|}$. Если это количество больше $|Y|/\sqrt{|X|}$, то объем памяти, равный $|X||Y|$, меньше $|S|\sqrt{|X|}$.

Так как алгоритм гарантирует размещение без склеиваний произвольного словаря объема S , то время поиска по первому ключу в худшем случае равно $|Y|$. Однако если у каждого первого ключа имеется по $|S|/|X|$ вторых ключей, то время поиска можно сократить до $|S|/|X|$, а поиск по второму ключу всегда требует $|X|$ шагов.

Для сравнения последнего алгоритма с предыдущими сведем в таблицу все параметры, вычисленные для списков, в которых имеется одинаковое число имен у любой фамилии.

Автор алгоритма	Время поиска				Объем памяти
	по первому ключу		по второму ключу		
	среднее	худшее	среднее	худшее	
Р. Райвест	$\sqrt{ S }$	$ S $	$\frac{ X Y }{\sqrt{ S }}$	$ S $	$ S $
Д. Слепян и Д. Вулф	$\frac{ S }{ X }$	$ S $	$ X $	$ S $	$ S ^{1+\varepsilon}$
М. Трофимова	—	$\frac{ S }{ X }$	—	$ X $	$ S \sqrt{ X }$

Замечание к таблице. На самом деле на пересечении первой строки с первым и третьим столбцами должны находиться числа $\max\{|S|/|X|, \sqrt{|S|}\}$ и $\max\{|X|, |X||Y|/\sqrt{|S|}\}$ соответственно.

Таким образом, время поиска с помощью предлагаемого алгоритма в наихудшем случае не превосходит среднего по словарям времени поиска методом Р. Райвеста и будет значительно меньше, когда размещаются списки, в которых доля имен, приходящихся на одну фамилию, от числа всех имен не превосходит $\min\{|Y|/|X|, |X|/|Y|\}$. В отличие от Д. Слепяна и Д. Вулфа в рассматриваемом случае возможен быстрый поиск для всех элементов списка. Такой поиск является целесообразным, когда с одинаковой вероятностью требуется найти любой элемент. Что касается объема памяти, то в предлагаемом алгоритме он больше, чем у Р. Райвеста, а также у Д. Слепяна и Д. Вулфа. Однако без допущения склеиваний его нельзя существенно уменьшить, не ухудшая времени поиска.

Работа состоит из двух частей. Во-первых, предложен простой алгоритм, позволяющий для любого подмножества S декартового произведения $X \times Y$ с одинаковым числом элементов из Y на каждый элемент из X предъявить две функции с областями значений не более $|X|$ и $|S|/\sqrt{|X|}$ соответственно, которые задают инъективное на S отображение. Во-вторых, показано, что не существует взаимно-однозначного отображения таких подмножеств в множество размера меньше чем $|S|\sqrt{|X|/2}$.

Суть алгоритма состоит в следующем: шаг первый — все фамилии и имена перенумеровываются отдельно; шаг второй — находятся два имени, для которых соответствующие им множества фамилий не пересекаются, и обоим именам дается один и тот же номер, а их множества фамилий объединяются (тем самым общее число номеров для имен уменьшается на единицу); шаг третий — возврат к шагу два, а если пар больше нет, то алгоритм заканчивает работу и результирующая нумерация является искомой.

§ 1. Теорема о существовании инъективного отображения

ОПРЕДЕЛЕНИЕ 1. Если S — подмножество декартова произведения произвольных множеств X и Y , то для любого x из X множество

$$Y_{S,x} = \{y \in Y \mid (x, y) \in S\}$$

назовем *проекцией* множества S в точке x на Y .

Аналогично определяется проекция $X_{S,y}$ множества S в каждой точке y из Y на множество X :

$$X_{S,y} = \{x \in X \mid (x, y) \in S\}.$$

ЗАМЕЧАНИЕ 1. Когда из контекста понятно, о проекции какого множества идет речь, будем опускать S и обозначать $Y_{S,x}$ и $X_{S,y}$ через Y_x и X_y соответственно.

ОПРЕДЕЛЕНИЕ 2. Подмножество S декартова произведения $X \times Y$ произвольных множеств X и Y назовем *равномерным* по X с параметром r , $r > 0$, если $|Y_x| = r$ при любом x из X .

Теорема 1. Пусть X и Y — произвольные множества, $|X| \geq 9$, а r — положительное целое. Тогда для каждого $S \subset X \times Y$ равномерного по X с параметром r найдутся целочисленные функции f и g такие, что

$$\begin{aligned} f: X &\mapsto [0, \dots, |X| - 1], \\ g: Y &\mapsto [0, \dots, \lfloor r\sqrt{|X|} \rfloor - 1], \end{aligned}$$

и пара (f, g) задает инъективное на S отображение множества $X \times Y$.

Лемма 1. Пусть X и Y — произвольные множества и S — такое подмножество из $X \times Y$, равномерное по X с параметром r , $r > 0$, что все проекции S на X — непустые. Тогда для каждого \dot{y} из Y такого, что $|X_{\dot{y}}| < (|Y| - 1)/(r - 1)$, найдется \ddot{y} из Y такое, что $X_{\dot{y}} \cap X_{\ddot{y}}$ — пустое множество.

ДОКАЗАТЕЛЬСТВО. Предположим противное. Пусть имеется $y_0 \in Y$ такое, что $|X_{y_0}| < (|Y| - 1)/(r - 1)$ и $X_{y_0} \cap X_y \neq \emptyset$ для любого $y \in Y \setminus \{y_0\}$. Таким образом, для любого $y \in Y \setminus \{y_0\}$ существует $x \in X_{y_0} \cap X_y$. Следовательно, по определению 1 $y \in Y_x$ для этого $x \in X_{y_0}$. Поэтому

$$Y \setminus \{y_0\} \subseteq \bigcup_{x \in X_{y_0}} (Y_x \setminus \{y_0\})$$

и

$$|Y \setminus \{y_0\}| \leq \sum_{x \in X_{y_0}} (|Y_x| - 1) = |X_{y_0}|(r - 1).$$

Значит, $|X_{y_0}| \geq (|Y| - 1)/(r - 1)$. Противоречие. Лемма 1 доказана.

ОПРЕДЕЛЕНИЕ 3. Пусть X и Y — произвольные множества и $P(X \times Y)$ — множество всех подмножеств декартова произведения $X \times Y$. Тогда для любого $S \in P(X \times Y)$ и любых $\dot{y} \in Y$, $\ddot{y} \in Y$ таких, что пересечение непустых множеств $X_{\dot{y}}$ и $X_{\ddot{y}}$ — пустое, определим преобразование «*»:

$$P(X \times Y) \times Y \times Y \mapsto P(X \times Y),$$

задаваемое формулой

$$S_{(\dot{y}, \ddot{y})}^* = (S \setminus \{(x, \dot{y}) \mid x \in X_{\dot{y}}\}) \cup \{(x, \ddot{y}) \mid x \in X_{\dot{y}}\}.$$

Лемма 2. Преобразование «*» сохраняет свойство равномерности множества с тем же параметром.

Доказательство. Пусть X и Y — произвольные множества, S — равномерное по X подмножество из $X \times Y$. Выберем такие $\dot{y}, \ddot{y} \in Y$, что $X_{S,\dot{y}} \neq \emptyset$ и $X_{S,\ddot{y}} \neq \emptyset$, а $X_{S,\dot{y}} \cap X_{S,\ddot{y}} = \emptyset$. Преобразование «*» применим к S и \dot{y}, \ddot{y} .

Чтобы убедиться в равномерности множества $S_{(\dot{y},\ddot{y})}^*$, согласно определению 2 для каждого $x \in X$ требуется вычислить мощность множества $Y_{S_{(\dot{y},\ddot{y})}^*,x}$. Воспользовавшись определениями 1 и 3, получаем

$$\begin{aligned} Y_{S_{(\dot{y},\ddot{y})}^*,x} &= \{y \in Y \mid (x, y) \in S_{(\dot{y},\ddot{y})}^*\} \\ &= \{y \in Y \mid (x, y) \in (S \setminus \{(x, \dot{y}) \mid x \in X_{S,\dot{y}}\}) \cup \{(x, \ddot{y}) \mid x \in X_{S,\ddot{y}}\}\} \\ &= (\{y \in Y \mid (x, y) \in S\} \setminus \{y \in Y \mid (x, y) \in \{(x, \dot{y}) \mid x \in X_{S,\dot{y}}\}\}) \\ &\quad \cup \{y \in Y \mid (x, y) \in \{(x, \ddot{y}) \mid x \in X_{S,\ddot{y}}\}\}. \end{aligned}$$

Следует заметить, что

$$\{y \in Y \mid (x, y) \in \{(x, \dot{y}) \mid x \in X_{S,\dot{y}}\}\} = \begin{cases} \{\dot{y}\}, & \text{если } x \in X_{S,\dot{y}}; \\ \emptyset & \text{в остальных случаях.} \end{cases}$$

Аналогично

$$\{y \in Y \mid (x, y) \in \{(x, \ddot{y}) \mid x \in X_{S,\ddot{y}}\}\} = \begin{cases} \{\ddot{y}\}, & \text{если } x \in X_{S,\ddot{y}}; \\ \emptyset & \text{в остальных случаях.} \end{cases}$$

Поэтому для любого $x \in X \setminus X_{S,\dot{y}}$ имеем

$$|Y_{S_{(\dot{y},\ddot{y})}^*,x}| = |\{y \in Y \mid (x, y) \in S\}| = |Y_{S,x}|.$$

Заметим, что если $x \in X_{S,\dot{y}}$, то $x \notin X_{S,\ddot{y}}$, так как $X_{S,\dot{y}} \cap X_{S,\ddot{y}} = \emptyset$. Следовательно, $\dot{y} \in Y_{S,x}$, $\ddot{y} \notin Y_{S,x}$ и

$$|Y_{S_{(\dot{y},\ddot{y})}^*,x}| = |(Y_{S,x} \setminus \{\dot{y}\}) \cup \{\ddot{y}\}| = (|Y_{S,x}| - 1) + 1 = |Y_{S,x}|.$$

Из доказанных равенств следует, что все проекции на Y множества $S_{(\dot{y},\ddot{y})}^*$ совпадают по объему с проекциями на Y множества S . Таким образом, равномерность S влечет равномерность $S_{(\dot{y},\ddot{y})}^*$ с тем же параметром. Лемма 2 доказана.

Лемма 3. Пусть X и Y — произвольные множества, f и g — функции, заданные на X и Y соответственно, $S \subseteq X \times Y$ и элементы \dot{y}, \ddot{y} из Y таковы, что пересечение непустых множеств $X_{S,\dot{y}}$ и $X_{S,\ddot{y}}$ — пустое. Тогда если отображение (f, g) инъективно на $S_{(\dot{y},\ddot{y})}^*$ и $g(\dot{y}) = g(\ddot{y})$, то отображение (f, g) инъективно на S .

Доказательство. Для краткости будем обозначать $S_{(\dot{y},\ddot{y})}^*$ через S^* . Докажем от противного. Предположим, что отображение (f, g) не является инъективным на S . Тогда существуют $(x_1, y_1) \in S$ и $(x_2, y_2) \in S$

такие, что $(f, g)(x_1, y_1) = (f, g)(x_2, y_2)$. По условию инъективности (f, g) на S^* возможен лишь один из следующих случаев:

- (a) $(x_1, y_1) \in S^*, (x_2, y_2) \notin S^*$;
- (b) $(x_1, y_1) \notin S^*, (x_2, y_2) \in S^*$;
- (c) $(x_1, y_1) \notin S^*, (x_2, y_2) \notin S^*$.

Согласно определению 3 S и S^* совпадают, за исключением подмножеств размерности $|X_{\dot{y}}|$, т. е.

$$S \setminus \{(x, \dot{y}) \mid x \in X_{\dot{y}}\} = S^* \setminus \{(x, \ddot{y}) \mid x \in X_{\ddot{y}}\}.$$

Следовательно, если $(x, y) \notin S^*$ и $(x, y) \in S$, то $y = \dot{y}$.

В случае (a) из равенств $y_2 = \dot{y}$ и $g(\dot{y}) = g(\ddot{y})$ следует, что

$$(f, g)(x_1, y_1) = (f, g)(x_2, y_2) = (f, g)(x_2, \dot{y}) = (f, g)(x_2, \ddot{y}),$$

причем $(x_1, y_1) \in S^*$ и $(x_2, \ddot{y}) \in S^*$, т. е. отображение (f, g) не инъективно на S^* . Противоречие.

Случай (b) рассматривается аналогично.

В случае (c) из равенств $y_1 = \dot{y}$, $y_2 = \dot{y}$ и $g(\dot{y}) = g(\ddot{y})$ следует, что

$$\begin{aligned} (f, g)(x_1, y_1) &= (f, g)(x_1, \dot{y}) = (f, g)(x_1, \ddot{y}), \\ (f, g)(x_2, y_2) &= (f, g)(x_2, \dot{y}) = (f, g)(x_2, \ddot{y}). \end{aligned}$$

Таким образом,

$$(f, g)(x_1, \ddot{y}) = (f, g)(x_1, y_1) = (f, g)(x_2, y_2) = (f, g)(x_2, \ddot{y}),$$

причем $(x_1, \ddot{y}) \in S^*$ и $(x_2, \ddot{y}) \in S^*$, т. е. отображение (f, g) не инъективно на S^* . Противоречие. Лемма 3 доказана.

ОПРЕДЕЛЕНИЕ 4. Если S — подмножество декартова произведения произвольных множеств X и Y , то для любого целого h , $0 \leq h \leq |X|$, определим множество

$$Y_S^{(h)} = \{y \in Y \mid |X_{S,y}| = h\}.$$

ЗАМЕЧАНИЕ 2. Если из контекста понятно, о каком подмножестве декартова произведения идет речь, будем опускать S и обозначать $Y_S^{(h)}$ через $Y^{(h)}$.

Лемма 4. Пусть X и Y — произвольные множества, S — подмножество из $X \times Y$, равномерное по X с параметром r , $r > 0$, все проекции которого на X — непустые. Тогда для любого h , $0 < h \leq |X|$, существует $T = T(S, h) \subseteq Y_S^{(h)} \times Y$, обладающее свойствами:

- 1) если $(\dot{y}, \ddot{y}) \in T$, то $(\ddot{y}, \dot{y}) \notin T$;
- 2) если для любых \dot{y}, \ddot{y} из $Z \subseteq Y$, $\dot{y} \neq \ddot{y}$, существует последовательность элементов $y_1, y_2, \dots, y_k \in Y$, $0 \leq k < |T|$, такая, что $(\dot{y}, y_1) \in T$, $(y_1, y_2) \in T, \dots, (y_k, \ddot{y}) \in T$, то для всех $y \in Z$ проекции $X_{S,y}$ множества S на X попарно не пересекаются;
- 3) $|T| \geq \min\{\frac{1}{2}|Y_S^{(h)}|, \max\{0, |Y| - h(r-1) - 1\}\}$.

Доказательство. Для произвольного h , $0 < h \leq |X|$, такого, что $Y_S^{(h)} \neq \emptyset$, наберем множество пар T следующим образом. Для $\dot{y}_1 \in Y_S^{(h)}$ выберем $\ddot{y}_1 \in Y$ такое, что $X_{S, \dot{y}_1} \cap X_{S, \ddot{y}_1} = \emptyset$.

Если $h < (|Y| - 1)/(r - 1)$, т. е. $|X_{S, \dot{y}_1}| < (|Y| - 1)/(r - 1)$, то \ddot{y}_1 существует по лемме 1. Добавим (\dot{y}_1, \ddot{y}_1) в T . Множество S преобразуем в множество $S_{(\dot{y}_1, \ddot{y}_1)}^*$. По лемме 2 $S_{(\dot{y}_1, \ddot{y}_1)}^*$ является равномерным по X с тем же параметром r . Пусть $\dot{y}_2 \in Y_S^{(h)} \setminus \{\dot{y}_1, \ddot{y}_1\} = Y_{S_{(\dot{y}_1, \ddot{y}_1)}^*}^{(h)}$ (равенство следует из определений 3 и 4). Найдем $\ddot{y}_2 \in \{y \in Y \mid X_{S_{(\dot{y}_1, \ddot{y}_1)}^*, y} \neq \emptyset\}$ такое, что $X_{S_{(\dot{y}_1, \ddot{y}_1)}^*, \dot{y}_2} \cap X_{S_{(\dot{y}_1, \ddot{y}_1)}^*, \ddot{y}_2} = \emptyset$. По лемме 1 такое \ddot{y}_2 можно найти, если $|\{y \in Y \mid X_{S_{(\dot{y}_1, \ddot{y}_1)}^*, y} \neq \emptyset\}| > h(r - 1) + 1$. Заметим, что по определению 3

$$\{y \in Y \mid X_{S_{(\dot{y}_1, \ddot{y}_1)}^*, y} \neq \emptyset\} = Y \setminus \{\dot{y}_1\}.$$

Следовательно, если $|Y| - 1 > h(r - 1) + 1$, то \ddot{y}_2 существует. Добавим (\dot{y}_2, \ddot{y}_2) в T .

К множеству $S_{(\dot{y}_1, \ddot{y}_1)}^*$ и паре (\dot{y}_2, \ddot{y}_2) применим преобразование «*» и найдем новый элемент множества T . Такую процедуру можно повторять, как минимум, пока справедливы оба неравенства:

$$Y_S^{(h)} \setminus \{\dot{y}_1, \ddot{y}_1, \dot{y}_2, \ddot{y}_2, \dots, \dot{y}_m, \ddot{y}_m\} \neq \emptyset, \\ |Y| - m > h(r - 1) + 1.$$

Таким образом,

$$|T| \geq \min \left\{ \frac{1}{2} |Y_S^{(h)}|, |Y| - h(r - 1) - 1 \right\},$$

если $|Y| - h(r - 1) - 1 > 0$.

Следовательно, для любого h , $0 < h \leq |X|$,

$$|T| \geq \min \left\{ \frac{1}{2} |Y_S^{(h)}|, \max\{0, |Y| - h(r - 1) - 1\} \right\}$$

и свойство 3 доказано.

Теперь проверим первые два свойства множества T .

Пусть $(\dot{y}, \ddot{y}) \in T$. Тогда на следующем шаге $\ddot{y} \notin Y_{S_{(\dot{y}, \ddot{y})}}^{(h)}$ и $\dot{y} \notin \{y \in Y \mid X_{S_{(\dot{y}, \ddot{y})}^*, y} \neq \emptyset\} = Y \setminus \{\dot{y}\}$. Поэтому $(\ddot{y}, \dot{y}) \notin T$, т. е. выполнено свойство 1.

Проверим справедливость свойства 2. Пусть $\dot{y}, \ddot{y} \in Z$, где $Z \subseteq Y$,

описанное в формулировке леммы. По построению множества T имеем

$$\begin{aligned} X_{\dot{y}} \cap X_{y_1} &= \emptyset, \\ X_{y_2} \cap (X_{\dot{y}} \cup X_{y_1}) &= \emptyset, \\ X_{y_3} \cap (X_{\dot{y}} \cup X_{y_1} \cup X_{y_2}) &= \emptyset, \\ &\dots \\ X_{\ddot{y}} \cap \left(X_{\dot{y}} \cup \left(\bigcup_{i=1}^k X_{y_i} \right) \right) &= \emptyset. \end{aligned}$$

Значит, множества $X_{\dot{y}}$ и $X_{\ddot{y}}$ не пересекаются.

Таким образом, построенное множество T удовлетворяет свойствам 1–3. Лемма 4 доказана.

Далее приводятся два алгоритма. Первый алгоритм сначала перенумеровывает элементы из $\{y \in Y \mid X_{S,y} \neq \emptyset\}$, потом строит множество $T(S, 1)$ и для каждой пары \dot{y}, \ddot{y} из этого множества элементу \dot{y} присваивает номер элемента \ddot{y} и т. д. Затем строится множество $T(W, 2)$, где W — множество, полученное в ходе построения $T(S, 1)$ из множества S , и снова производится смена номеров. Алгоритм продолжаем строить $T(W, h)$, пока это возможно. Второй алгоритм используется для удаления номеров, которые освободились в ходе первого алгоритма.

Алгоритм АВ.

Пусть X и Y — произвольные множества, $S \subset X \times Y$, α — целочисленная функция, h — целочисленная переменная, W — текущее множество и V — последнее значение текущего множества W .

Шаг 0. Элементы из $\{y \in Y \mid X_{S,y} \neq \emptyset\}$ перенумеровываются с помощью функции α .

Шаг 1. $h = 0$, $W = S$.

Шаг 2. $h = h + 1$.

Шаг 3. Пусть $\dot{y} \in Y_W^{(h)}$. Если существует $\ddot{y} \in Y$ такое, что $X_{W,\dot{y}} \neq \emptyset$, а $X_{W,\dot{y}} \cap X_{W,\ddot{y}} = \emptyset$, то шаг 4. Иначе шаг 7.

Шаг 4. $\alpha(\dot{y}) = \alpha(\ddot{y})$.

Шаг 5. Преобразование «*» применяется к W и (\dot{y}, \ddot{y}) , и полагается $W = W_{(\dot{y}, \ddot{y})}^*$.

Шаг 6. Если $Y_W^{(h)} \neq \emptyset$, то шаг 3, иначе шаг 7.

Шаг 7. Если $h < |X|$, то шаг 2.

Шаг 8. $V = W$.

Конец.

Алгоритм CD.

Пусть K — положительное целое, α и β — целочисленные функции, k — целочисленная переменная.

Шаг 1. $\beta(0) = -1, k = 0$.

Шаг 2. $k = k + 1$.

Шаг 3. Если $\alpha^{-1}(k) = \emptyset$, то $\beta(k) = \beta(k - 1)$.

Шаг 4. Если $\alpha^{-1}(k) \neq \emptyset$, то $\beta(k) = \beta(k - 1) + 1$.

Шаг 5. Если $k < K$, то шаг 2.

Конец.

§ 2. Доказательство теоремы 1

Пусть S — подмножество в $X \times Y$, равномерное по X с параметром r , f — заданная на X композиция отображений β_1 и α_1 , где α_1 — произвольная нумерация множества X , а β_1 — отображение, полученное с помощью алгоритма CD при $K = |X|$ и $\alpha = \alpha_1$. Положим $g = \beta_2 \circ \alpha_2$, где α_2 — заданное на $\{y \in Y \mid X_{S,y} \neq \emptyset\}$ отображение, полученное по алгоритму AB из произвольной нумерации этого множества, а β_2 — отображение, полученное с помощью алгоритма CD при $K = |Y|$ и $\alpha = \alpha_2$. Докажем, что, во-первых, (f, g) — инъективное на S отображение, во-вторых, размер области значений функции g не превосходит $\lfloor r\sqrt{|X|} \rfloor$. Тогда теорема будет доказана, так как размер области значений функции f не превосходит $|X|$.

Инъективность докажем индукцией по значениям множества W (из алгоритма AB).

Покажем, что отображение (f, g) инъективно на V — последнем значении множества W . Функция f инъективна на X по построению. Функция g является композицией двух инъективных функций: β_2 инъективна на $\{k \mid \alpha_2^{-1}(k) \neq \emptyset\}$ по построению и α_2 инъективна на $\{y \in Y \mid X_{V,y} \neq \emptyset\}$, так как ее значения на этом множестве совпадают с первоначальной нумерацией на шаге 0 в алгоритме AB . Поэтому отображение (f, g) инъективно на любом подмножестве из $X \times \{y \in Y \mid X_{V,y} \neq \emptyset\}$, в том числе на V .

Заметим, что текущее множество W является результатом преобразования «*», которое применили к предыдущему значению множества W и паре (\dot{y}, \ddot{y}) . Значит, если (f, g) инъективно на текущем значении множества W и по шагу 4 алгоритма AB $g(\dot{y}) = g(\ddot{y})$, то согласно лемме 3 отображение (f, g) инъективно на предыдущем значении множества W . Таким образом, дойдем до исходного множества S и получим, что (f, g) — инъективное на S отображение.

Докажем, что размер области значений функции g не превосходит $\lfloor r\sqrt{|X|} \rfloor$. Этот размер равен

$$\begin{aligned} |\{k \mid \alpha_2^{-1}(k) \neq \emptyset\}| &= |\{y \in Y \mid X_{V,y} \neq \emptyset\}| \\ &= |Y| - |\{y \in Y \mid X_{V,y} = \emptyset\}| = |Y| - |Y_V^{(0)}|, \end{aligned}$$

где V из алгоритма AB .

Если $|\{y \in Y \mid X_{S,y} \neq \emptyset\}| \leq \lfloor r\sqrt{|X|} \rfloor$, то утверждение теоремы очевидно.

В случае, когда $|\{y \in Y \mid X_{S,y} \neq \emptyset\}| > \lfloor r\sqrt{|X|} \rfloor$, теорему будем доказывать от противного. Пусть

$$|Y| - |Y_V^{(0)}| > \lfloor r\sqrt{|X|} \rfloor.$$

По построению V — равномерное по X с параметром r подмножество в $X \times (Y \setminus Y_V^{(0)})$ и все проекции V на X — непустые. Так как V — последнее значение W из алгоритма AB , то нельзя подобрать пару ни для одного $y \in Y \setminus Y_V^{(0)}$. Значит, для любого h , $0 < h \leq |X|$, определенное в лемме 4 множество $T = T(V, h) \subseteq Y_V^{(h)} \times (Y \setminus Y_V^{(0)})$ является пустым. Следовательно, по лемме 4 $Y_V^{(h)} = \emptyset$ для всех $h < (|Y| - |Y_V^{(0)}| - 1)/(r - 1)$. Поскольку $V \neq \emptyset$, существует h_0 такое, что

$$h_0 = \min\{h > 0 \mid Y_V^{(h)} \neq \emptyset\}.$$

Тогда из леммы 4 следует, что

$$h_0 \geq (|Y| - |Y_V^{(0)}| - 1)/(r - 1).$$

В свою очередь,

$$|S| = |X|r = \sum_{y \in Y} |X_{V,y}| \geq \sum_{y \in Y \setminus Y_V^{(0)}} h_0 = h_0(|Y| - |Y_V^{(0)}|).$$

Используя нижние оценки для h_0 и $(|Y| - |Y_V^{(0)}|)$, получаем

$$|X|r > \lfloor \sqrt{|X|}r \rfloor (\lfloor \sqrt{|X|}r \rfloor - 1)/(r - 1),$$

$$|X|r(r - 1) > \lfloor \sqrt{|X|}r \rfloor^2 - \lfloor \sqrt{|X|}r \rfloor,$$

$$|X|r(r - 1) > (\sqrt{|X|}r - 1)^2 - \sqrt{|X|}r,$$

$$|X|r(r - 1) > |X|r^2 - 3\sqrt{|X|}r + 1.$$

Следовательно, $|X|r + 1 < 3\sqrt{|X|}r$. Это противоречит условию $|X| \geq 9$. Теорема 1 доказана.

§ 3. Обратная теорема

Теорема 2. Пусть X и Y — произвольные множества, а r — положительное целое. Тогда в $X \times Y$ найдется такое подмножество S , равномерное по X с параметром r , что если $|Y| \geq r\sqrt{|X|}/2$ и целое $M < \lfloor r/2 \rfloor \lfloor \sqrt{2|X|} \rfloor$, то не существует целочисленных функций

$$f: X \mapsto [0, \dots, |X| - 1],$$

$$g: Y \mapsto [0, \dots, M - 1],$$

задающих инъективное на S отображение.

Лемма 5. Пусть X и Y — произвольные множества, а r — положительное целое. Тогда в $X \times Y$ найдется подмножество S такое, что любая его проекция на Y имеет мощность r , а все его непустые проекции на X попарно пересекаются, причем

$$|\{y \in Y \mid X_y \neq \emptyset\}| \geq \begin{cases} \lfloor r/2 \rfloor \lfloor \frac{2|Y|}{r} \rfloor, & \text{если } |Y| < r\sqrt{\frac{|X|}{2}}; \\ \lfloor r/2 \rfloor \lfloor \sqrt{2|X|} \rfloor, & \text{если } |Y| \geq r\sqrt{\frac{|X|}{2}}. \end{cases}$$

Доказательство. Пусть $|Y| \geq r\sqrt{|X|/2}$ и r — четное. Положим $k = \lfloor \sqrt{2|X|} \rfloor$ и выберем k произвольных непересекающихся подмножеств Y^1, \dots, Y^k из Y таких, что $|Y^i| = r/2$, $1 \leq i \leq k$. Из этих подмножеств составим набор всевозможных неупорядоченных пар

$$Z = \{Z^{i,j} = Y^i \cup Y^j \mid i = 1, \dots, k; j = i+1, \dots, k\}.$$

Ясно, что $|Z| = k(k-1)/2 \leq |X|$ и $|Z^{i,j}| = r$ для любых i и j , $1 \leq i \leq k$, $1 \leq j \leq k$, $i \neq j$.

Возьмем произвольное однозначное отображение ω из X на Z . Построим $S \subseteq X \times Y$ с помощью отображения ω :

$$S = \{(x, y) \mid x \in X, y \in \omega(x)\}.$$

Установим следующие свойства множества S . Во-первых, для всех $x \in X$ проекция $Y_x = \omega(x) \in Z$; следовательно, $|Y_x| = r$ по определению множества Z .

Во-вторых, для любых $\dot{y}, \ddot{y} \in \{y \in Y \mid X_y \neq \emptyset\}$ справедливо одно из утверждений:

- (а) существует i такое, что $\dot{y} \in Y^i$ и $\ddot{y} \in Y^i$;
- (б) существуют i и j такие, что $\dot{y} \in Y^i$, а $\ddot{y} \in Y^j$.

Воспользуемся следующими двумя фактами. По построению множества S его проекция на X в любой точке $\dot{y} \in Y^i$ для каждого i , $1 \leq i \leq k$,

$$\begin{aligned} X_{\dot{y}} &= \{x \in X \mid \dot{y} \in \omega(x)\} = \{x \in X \mid Y^i \subset \omega(x)\} \\ &= \bigcup_{l=1}^{i-1} \{x \in X \mid \omega(x) = Z^{i,l}\} \bigcup_{l=i+1}^k \{x \in X \mid \omega(x) = Z^{i,l}\} \\ &= \bigcup_{l=1, l \neq i}^k \omega^{-1}(Z^{i,l}). \end{aligned}$$

Из однозначности отображения ω следует, что отображение ω^{-1} — разноточное, т. е.

$$\omega^{-1}(Z^{l,m}) \cap \omega^{-1}(Z^{s,t}) = \begin{cases} \omega^{-1}(Z^{l,m}), & \text{если } Z^{l,m} = Z^{s,t}; \\ \emptyset, & \text{если } Z^{l,m} \neq Z^{s,t}. \end{cases}$$

В случае (а) имеем

$$X_{\dot{y}} = \{x \in X \mid \dot{y} \in \omega(x)\} = \{x \in X \mid Y^i \subseteq \omega(x)\} = X_{\ddot{y}}.$$

Значит, $X_{\dot{y}} \cap X_{\ddot{y}} = X_{\dot{y}} = X_{\ddot{y}} \neq \emptyset$.

В случае (б) имеем

$$\begin{aligned} X_{\dot{y}} \cap X_{\ddot{y}} &= \left(\bigcup_{l=1, l \neq i}^k \omega^{-1}(Z^{i,l}) \right) \cap \left(\bigcup_{m=1, m \neq j}^k \omega^{-1}(Z^{j,m}) \right) \\ &= \bigcup_{l=1, l \neq i}^k \bigcup_{m=1, m \neq j}^k \left(\omega^{-1}(Z^{i,l}) \cap \omega^{-1}(Z^{j,m}) \right) = \omega^{-1}(Z^{i,j}) \neq \emptyset, \end{aligned}$$

т. е. при $|Y| \geq r\sqrt{|X|/2}$ и четном r лемма доказана.

Пусть $|Y| \geq r\sqrt{|X|/2}$ и r — нечетное, $r = 2p + 1$. Тогда доказательство проведем по той же схеме, что и раньше. Положим $k = \lfloor \sqrt{2|X|} \rfloor$ и выберем k произвольных непересекающихся p -элементных подмножеств Y^1, \dots, Y^k из Y . Из этих подмножеств составим всевозможные неупорядоченные пары, выделим элемент $y^* \in Y \setminus \left(\bigcup_{i=1}^k Y^i \right)$ и определим множество

$$Z = \{Z^{i,j} = Y^i \cup Y^j \cup \{y^*\} \mid i = 1, \dots, k; j = i + 1, \dots, k\}.$$

Ясно, что $|Z| = k(k-1)/2 \leq |X|$ и $|Z^{i,j}| = r$ для любых i и j , $1 \leq i \leq k$, $1 \leq j \leq k$, $i \neq j$.

Возьмем произвольное однозначное отображение ω из X на Z . Построим $S \subseteq X \times Y$ с помощью отображения ω :

$$S = \{(x, y) \mid x \in X, y \in \omega(x)\}.$$

Установим следующие свойства множества S . Во-первых, для всех $x \in X$ проекция $Y_x = \omega(x) \in Z$. Следовательно, $|Y_x| = r$ по определению множества Z .

Во-вторых, для любых $\dot{y}, \ddot{y} \in \{y \in Y \mid X_y \neq \emptyset\}$ справедливо одно из утверждений:

- (а) существует i такое, что $\dot{y} \in Y^i$ и $\ddot{y} \in Y^i$;
- (б) существуют i и j такие, что $\dot{y} \in Y^i$, а $\ddot{y} \in Y^j$;
- (с) существует i такое, что $\dot{y} \in Y^i$, а $\ddot{y} = y^*$.

В случае (а) по построению множества S имеем

$$X_{\dot{y}} = \bigcup_{l=1, l \neq i}^k \omega^{-1}(Z^{i,l}) = X_{\ddot{y}}.$$

Значит, $X_{\dot{y}} \cap X_{\ddot{y}} = X_{\dot{y}} = X_{\ddot{y}} \neq \emptyset$.

В случае (b) из однозначности отображения ω следует, что

$$\begin{aligned} X_{\dot{y}} \cap X_{\ddot{y}} &= \left(\bigcup_{l=1, l \neq i}^k \omega^{-1}(Z^{i,l}) \right) \cap \left(\bigcup_{m=1, m \neq j}^k \omega^{-1}(Z^{j,m}) \right) \\ &= \bigcup_{l=1, l \neq i}^k \bigcup_{m=1, m \neq j}^k \left(\omega^{-1}(Z^{i,l}) \cap \omega^{-1}(Z^{j,m}) \right) = \omega^{-1}(Z^{i,j}) \neq \emptyset. \end{aligned}$$

В случае (c) по определению множества Z имеем

$$X_{\dot{y}} \cap X_{y^*} = X_{\dot{y}} \cap X \neq \emptyset,$$

т. е. при $|Y| \geq r\sqrt{|X|/2}$ и нечетном r лемма доказана.

Пусть $|Y| < r\sqrt{|X|/2}$. Положим $k = \lfloor (2/r)|Y| \rfloor$ и выберем k произвольных непересекающихся подмножеств Y^1, \dots, Y^k из Y таких, чтобы каждое подмножество имело объем $\lfloor r/2 \rfloor$. Дальнейшие рассуждения аналогичны случаю, когда $|Y| \geq r\sqrt{|X|/2}$.

Осталось заметить, что при четном r и $|Y| \geq r\sqrt{|X|/2}$

$$|\{y \in Y \mid X_y \neq \emptyset\}| = \left| \bigcup_{i=1}^k Y^i \right| = \sum_{i=1}^k |Y^i| = kr/2 \geq \lfloor \sqrt{2|X|} \rfloor r/2;$$

при нечетном r и $|Y| \geq r\sqrt{|X|/2}$

$$|\{y \in Y \mid X_y \neq \emptyset\}| = \sum_{i=1}^k |Y^i| + 1 = k(r-1)/2 + 1 \geq \lfloor r/2 \rfloor \lfloor \sqrt{2|X|} \rfloor;$$

при четном r и $|Y| < r\sqrt{|X|/2}$

$$|\{y \in Y \mid X_y \neq \emptyset\}| = kr/2 \geq \lfloor 2|Y|/r \rfloor r/2;$$

при нечетном r и $|Y| < r\sqrt{|X|/2}$

$$|\{y \in Y \mid X_y \neq \emptyset\}| = k(r-1)/2 + 1 \geq \lfloor r/2 \rfloor \lfloor 2|Y|/r \rfloor.$$

Следовательно,

$$|\{y \in Y \mid X_y \neq \emptyset\}| \geq \begin{cases} \lfloor r/2 \rfloor \lfloor \frac{2|Y|}{r} \rfloor, & \text{если } |Y| < r\sqrt{\frac{|X|}{2}}; \\ \lfloor r/2 \rfloor \lfloor \sqrt{2|X|} \rfloor, & \text{если } |Y| \geq r\sqrt{\frac{|X|}{2}}. \end{cases}$$

Лемма 5 доказана.

Лемма 6. Пусть X и Y — произвольные множества, S — подмножество в $X \times Y$ такое, что все его непустые проекции на X попарно пересекаются. Тогда если целое $M < |\{y \in Y \mid X_y \neq \emptyset\}|$, то при любом целом $N > 0$ не существует функций f и g таких, что

$$\begin{aligned} f : X &\mapsto [0, \dots, N-1], \\ g : Y &\mapsto [0, \dots, M-1], \end{aligned}$$

и которые задавали бы отображение множества $X \times Y$, инъективное на S .

ДОКАЗАТЕЛЬСТВО. Докажем от противного. Пусть существует пара функций (f, g) , которая задает инъективное на S отображение, причем $M < |\{y \in Y \mid X_y \neq \emptyset\}|$. Тогда существуют $\dot{y} \in Y$ и $\ddot{y} \in Y$ такие, что $X_{\dot{y}} \neq \emptyset$, $X_{\ddot{y}} \neq \emptyset$ и $g(\dot{y}) = g(\ddot{y})$. Из свойства множества S следует, что $X_{\dot{y}} \cap X_{\ddot{y}} \neq \emptyset$. Значит, найдется $x \in X_{\dot{y}} \cap X_{\ddot{y}}$, т. е. $(x, \dot{y}) \in S$ и $(x, \ddot{y}) \in S$. В свою очередь,

$$(f, g)(x, \dot{y}) = (f(x), g(\dot{y})) = (f(x), g(\ddot{y})) = (f, g)(x, \ddot{y}),$$

т. е. отображение (f, g) не инъективно на S . Противоречие. Лемма 6 доказана.

ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 2. По лемме 5 существует подмножество S из $X \times Y$ такое, что, во-первых, $|Y_x| = r$ при любом $x \in X$, во-вторых, все непустые проекции S на X попарно пересекаются, причем

$$|\{y \in Y \mid X_y \neq \emptyset\}| \geq \begin{cases} \lfloor r/2 \rfloor \lfloor \frac{2|Y|}{r} \rfloor, & \text{если } |Y| < r\sqrt{\frac{|X|}{2}}; \\ \lfloor r/2 \rfloor \lfloor \sqrt{2|X|} \rfloor, & \text{если } |Y| \geq r\sqrt{\frac{|X|}{2}}. \end{cases}$$

По лемме 6 для такого подмножества S нельзя подобрать пару функций (f, g) , инъективную на S , если область значений функции g меньше чем $|\{y \in Y \mid X_y \neq \emptyset\}|$. Так как по условию теоремы $|Y| > r\sqrt{|X|/2}$, то область значений функции g не может быть меньше чем $\lfloor r/2 \rfloor \lfloor \sqrt{2|X|} \rfloor$, что и требовалось доказать. Теорема 2 доказана.

Автор выражает благодарность Р. Е. Кричевскому, под руководством которого выполнена настоящая работа, за постановку задачи и ценные замечания, а также В. Н. Потапову за полезное обсуждение результатов работы.

ЛИТЕРАТУРА

1. **Кнут Д.** Искусство программирования для ЭВМ. М: Мир, 1978.
2. **Колесник В. Д., Полтырев Г. Ш.** Курс теории информации. М: Наука, 1982.
3. **Кричевский Р. Е.** Сжатие и поиск информации. М: Радио и связь, 1989.
4. **Slepian D., Wolf J.** Noiseless coding of correlated information sources // IEEE Trans. Inform. Theory. 1973. V. IT-19, N 4. P. 471-480.

Адрес автора:

Россия,
630090 Новосибирск,
ул. Пирогова, 2,
Новосибирский
государственный университет

Статья поступила

9 октября 1996 г.