УДК 519.2

Я. ДЕМЕТРОВИЧ, Д. КАТОНА

ЭКСТРЕМАЛЬНЫЕ ПРОБЛЕМЫ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ

ВВЕДЕНИЕ

Для пользователя важной характеристикой структуры базы данных (БД) является наличие соотношений между хранимыми в ней данными. Структуру БД образуют связи между типами записей, поэтому модели данных классифицируем по определению связи, допустимой между типами записей. На основе этого различаем две модели данных: сетевую и реляционную. В реляционных БД связь между множеством свойств БД осуществляется с помощью типа записи.

Задание типа записи — это определение «области определения» и «области значения». Заданный таким образом тип записи представляет собой функцию, определенную на области определений и принимающую значения в области значений.

При проектировании БД после задания ее структуры необходимо описать условия, которым БД должна удовлетворять в любой момент времени.

В реляционных БД одним из важных классов условий является функциональная зависимость, определенная в [1]. Содержимое реляционной БД можно представить как матрицу, поэтому функциональную зависимость определим для матриц. Если A и B — подмножества области определения матрицы M (т. е. множество атрибутов), то можно говорить, что B функционально зависит от A в матрице M, если любые две такие строки M, которые на любом элементе A принимают одинаковые значения, совпадают и на всех элементах B.

Проблемы, связанные с функциональной зависимостью, можно разделить на две группы. Первая — при проектировании реляционной БД условия, заданные пользователем для будущей БД, обычно не полны, т. е. некоторые, еще не сформулированные условия вытекают из заданных. Знание этих условий часто важно для пользователя, поэтому необходима разработка алгоритмов, с помощью которых можно найти условия, вытекающие из заданной системы условий.

В настоящей статье будут рассмотрены проблемы, входящие во вторую группу: комбинаторная сложность функциональных зависимостей данной матрицы.

В разд. 1 исследована структура системы минимальных ключей, представляющей собой систему Спернера. Показано, что любая система Спернера представляется матрицей как система минимальных ключей.

В разд. 2 дана оценка числа строк реляции, представляющей данную систему Спернера.

В разд. 3 определено максимальное число «существенных» функциональных зависимостей.

1. МАКСИМАЛЬНОЕ ЧИСЛО МИНИМАЛЬНЫХ КЛЮЧЕЙ

Пусть матрица M состоит из разных строк и содержит целые неотрицательные значения. Подмножество столбцов K назовем ключом, если строки подматрицы, определяемые этими столбцами, разные. Это условие означает, что одну строку определяют значения, стоящие в этих столбцах. Иначе говоря, если необходимо найти какую-то строку, достаточно задать значения, которые стоят в столбцах, относящихся к K.

Если K_1 — ключ и $K_1 \subset K_2$, то тогда, очевидно, K_2 также является ключом. Следовательно, представляют интерес только те ключи, для которых нет «меньшего» ключа. Те ключи K, для которых нет ключа $K' \subsetneqq K$, назовем минимальными, их множество обозначим $\mathcal{K}$.

В модели имеются четыре параметра: 1) число столбцов n; 2) число строк m; 3) максимальное число N, стоящее в матрице; 4) число минимальных ключей $|\mathcal{K}|$. Теперь сформулируем в общем виде проблему: как, зная три из приведенных параметра, определить пределы, в которых находится четвертый. Раскроем связь между n и $|\mathcal{K}|$. Следующая теорема сводит эту проблему к известной теореме, относящейся к системам множеств.

Теорема 1 [2]. Для системы подмножеств $\mathcal{K}$ множества, состоящей из n элементов, можно найти матрицу $n \times m$, в которой система минимальных ключей совпадает с $\mathcal{K}$ тогда и только тогда, когда для элементов $\mathcal{K}$ выполняется следующее условие:

$$\forall\, i, j \;\; K_i, K_j \in \mathcal{K}, \quad K \neq K_j \Rightarrow K_i \not\subset K_j, \tag{1}$$

т. е. $\mathcal{K}$ — система Спернера.

Доказательство. Необходимость условия (1) следует из определения минимальных ключей. Для доказательства достаточности предположим, что $\mathcal{K}$ удовлетворяет условию (1), и сконструируем матрицу. Введем следующую систему множеств: $\mathcal{K}^{-1} = \{B : B$ не содержит подмножество из $\mathcal{K}$ и является максимальным среди всех множеств с таким свойством$\}$.

Пусть

$$\mathcal{K}^{-1} = \{B_1, B_2, \ldots, B_m\}. \tag{2}$$

У искомой матрицы M будет $m+1$ строк. Нулевая строка состоит целиком из нулей, j-й элемент i-й строки ($1 \leqslant i \leqslant m$) будет 0 или i в зависимости от того, $j \in B_i$ или $j \notin B_i$. Ясно, что если $A \subseteq \{1, 2, \ldots, n\}$ —такое множество, которое не содержит подмножество из $\mathcal{K}$, то существует такое i, что $A \subseteq B_i$ и на A нулевая и i-я строки совпадают, поэтому A не является ключом. Если $K \in \mathcal{K}$, то тогда для любого i, $1 \leqslant i \leqslant m$, $K \setminus B_i \neq \varnothing$ и в каком-то столбце из K на i-й строке находится i, поэтому K будет ключом. Таким образом, система минимальных ключей M совпадает с $\mathcal{K}$.

Теорема 1 доказана.

В качестве иллюстрации рассмотрим следующий пример. Пусть $n = 5$, $\mathcal{K} = \{\{1, 2, 3\}, \{1, 2, 4\}, \{3, 4\}, \{4, 5\}\}$. Тогда $\mathcal{K}^{-1}\{\{1, 4\}, \{2, 4\}, \{1, 3, 5\}, \{2, 3, 5\}, \{1, 2, 5\}\}$ и соответствующей матрицей будет следующая:

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 2 & 0 & 2 & 0 & 2 \\ 0 & 3 & 0 & 3 & 0 \\ 4 & 0 & 0 & 4 & 0 \\ 0 & 0 & 5 & 5 & 0 \end{matrix}. \qquad (3)$$

Исходной задачей было определение верхней границы числа минимальных ключей матрицы, состоящей из n столбцов (нижняя граница не представляет интереса). На основании теоремы 1 можно сформулировать следующую задачу: сколько не содержащих друг друга подмножеств имеет множество из n элементов?

На этот вопрос дает ответ теорема Спернера [3], по которой этот максимум равен $\binom{n}{\left[\frac{n}{2}\right]}$ ($[x]$ означает целую часть x).

Теорема 2 [2]. В матрице M из n столбцов имеется не более чем $\binom{n}{\left[\frac{n}{2}\right]}$ минимальных ключей.

2. МИНИМАЛЬНОЕ ЧИСЛО СТРОК К ДАННЫМ КЛЮЧАМ

По теореме 1 для любой системы Спернера $\mathcal{K}$ можно найти такую матрицу M, у которой система минимальных ключей совпадает с $\mathcal{K}$. Однако минимальное число строк не выяснялось. Пусть $s(\mathcal{K})$—минимальное число строк. Подсчитаем $s(\mathcal{K})$ в случае разных систем Спернера. Доказательство теоремы 1 дает оценку на $s(\mathcal{K})$. В силу максимальности элементов $\mathcal{K}^{-1}$ следует, что $\mathcal{K}^{-1}$ является системой Спернера. Поэтому на основе теоремы Спернера можно сформулировать следующую теорему.

Теорема 3. Предположим, что

$$s(\mathcal{K}) \leqslant 1 + |\mathcal{K}^{-1}| \leqslant 1 + \binom{n}{\left[\frac{n}{2}\right]}. \qquad (4)$$

Следующая теорема покажет, что существует такое $\mathcal{K}$, для которого $s(\mathcal{K})$ «подходит близко» к верхней границе.

Теорема 4 [4]. При любом натуральном числе n можно найти такую систему Спернера $\mathcal{K}$ на n элементах, для которой

$$\frac{1}{n^2}\binom{n}{\left[\frac{n}{2}\right]} \leqslant s(\mathcal{K}). \qquad (5)$$

Доказательство. Пусть $s(n) = \max_{\mathcal{K}} \{s(\mathcal{K})\}$, где максимум берется по системам Спернера, построенным на n элементах. Матрица M реализует $\mathcal{K}$, если системой минимальных ключей M является $\mathcal{K}$. По приведенным определениям $\mathcal{K}$ на любых n элементах реализуется матрицей $m \times n$, где $m \leqslant s(n)$. Если в $(m + 1)$-ю, $(m + 2)$-ю, ..., $s(n)$-ю строки записать везде $(m + 1)$, $(m + 2)$, ..., $s(n)$, то тогда и расширенная матрица реализует $\mathcal{K}$. Таким образом, любое $\mathcal{K}$ реализуется матрицей $s(n) \times n$. Можно предположить, что в исходной матрице имеются только 1, 2, ..., m.

Рассмотрим элементы, стоящие в первом столбце M и упорядоченные по величине: $i_1 < i_2 < \dots$ $\dots i_r$ $(1 \leqslant r \leqslant s(n))$, заменим i_j на j. Можно легко убедиться, что это не меняет системы минимальных ключей. Сделав те же изменения во всех столбцах, придем к такому M, в котором все элементы $1, \dots, s(n)$. Число таких матриц $s(n)^{ns(n)}$. Поэтому

$$s(n)^{ns(n)} \geqslant \| \mathcal{K} \|, \qquad (6)$$

где $\| \mathcal{K} \|$ — мощность всех различных $\mathcal{K}$.

Поскольку система, состоящая из некоторых $\left[\frac{n}{2}\right]$-элементных множеств, всегда представляет собой систему Спернера, после уменьшения правой стороны в (6) придем к неравенству

$$s(n)^{ns(n)} > 2^{\binom{n}{\left]\frac{n}{2}\right[}}. \qquad (7)$$

Преобразуя (7), получим неравенство

$$s(n) \log_2 s(n) > \frac{1}{n}\binom{n}{\left[\frac{n}{2}\right]}. \qquad (8)$$

Если в (8) вместо $s(n)$ записать $\frac{1}{n^2}\binom{n}{\left[\frac{n}{2}\right]}$, получим величину, меньшую, чем в (8), т. е. из (8) следует, что $s(n) > \frac{1}{n^2}\binom{n}{\left[\frac{n}{2}\right]}$, что и доказывает справедливость теоремы 4.

Определим далее $s(\mathcal{K})$ для некоторых простых систем Спернера $\mathcal{K}$. Пусть F_k^n — система всех n-элементных подмножеств множества, состоящего из n элементов. Очевидно, что F_k^n является системой Спернера.

Лемма 1 [5]:

$$\left(s\left(\begin{matrix} F_k^n \\ 2 \end{matrix}\right)\right) \geqslant \left(\begin{matrix} n \\ k-1 \end{matrix}\right) \quad (0 \leqslant k \leqslant n).$$

Доказательство. Предположим, что некоторая матрица M, $m \times n$, реализует F_k^n. К любым разным $(k-1)$ столбцам можно найти пару строк, которые в этих столбцах совпадают. К другой системе — $(k-1)$ столбцов должна относиться другая пара строк. Поэтому $\binom{m}{2} \geqslant \binom{n}{k-1}$.

Лемма 1 доказана.

Из этой леммы следует, что $s(F_1^n) \geqslant 2$. Следующая конструкция показывает, что $s(F_1^n) = 2$:

$$\begin{matrix} 0 & 0 \ldots 0 \\ 1 & 1 \ldots 1 \end{matrix}$$

Определим теперь $s(F_2^n)$. Если матрица $m \times n$ реализует F_2^n, то по лемме 1

$$n \leqslant \binom{m}{2}. \qquad (9)$$

Если (9) выполняется, то можно построить матрицу $m \times n$, которая реализует F_2^n. Каждый столбец M содержит два нуля, и в разных столбцах пара нулей стоит на разных местах. Остальные элементы i-й строки будут i. Можно убедиться, что в одном столбце имеются две одинаковые строки, в то время как в двух — уже нет.

Следовательно, $s(F_2^n)$ — самое целое число, удовлетворяющее (9) в m.

Применим теперь лемму 1 для $k = n - 1$, $s(F_{n-1}^n) \geqslant n$. Следующая конструкция обеспечивает равенство:

$$\begin{matrix} 1 & 0 \ldots 0 \\ 0 & 1 \ldots 0 \\ \vdots & \\ 0 & 0 \ldots 1 \end{matrix}$$

Для определения $s(F_n^n)$ необходима еще одна лемма. Если M — какая-то матрица $m \times n$, то $G(M)$ будет графом, узлами которого являются строки M; два узла соединены, если две строки совпадают хотя бы в одном месте. К ребрам этого графа припишем множество общих мест A.

Лемма 2 [5]. Пусть M — матрицы и пусть по какому-то циклу графа $G(M)$ заданы множества $A_1, \ldots$ $\ldots, A_r$. Тогда

$$\text{для любого } 1 \leqslant j \leqslant r \bigcap_{\substack{i=1 \\ i=j}}^{r} A_i \setminus A_j = \varnothing. \qquad (10)$$

Доказательство. Предположим от противного, что (10) непусто, иначе говоря, что существует столбец, например u-й, который содержится в каждом A_i, кроме A_j. Пусть $k_1, \ldots, k_r$ — узлы, находящиеся в цикле так, что на (k_i, k_{i+1}) написан A_i $(1 \leqslant i < r)$ и на (k_r, k_1) — A_r. Из $u \in A_{j+1}$ следует, что k_{j+1}-е и k_{j+2}-е значения в u-й строке совпадают. То же самое можно сказать и о значениях k_{j+2}-й и k_{j+3}-й строк и т. д. Следовательно, k_{j+1}-е, k_{j+2}-е, ..., k_r-е, k_1-е, ..., k_j-е значения в u-м столбце совпадают. Из этого следует $u \in A_j$, что находится в противоречии с предположением.

Лемма 2 доказана.

Теперь можно определить $s(F_n^n)$. Предположим, что матрица M, $m \times n$, реализует F_n^n. Поскольку множество, состоящее из $(n-1)$ столбцов, не является ключом, оно должно быть приписано некоторому ребру $G(M)$. Следовательно, у графа $G(M)$ имеется n разных ребер, к которым приписаны множества из $(n-1)$ элементов. Эти ребра по лемме 2 не могут образовывать цикл, потому что совокупность множеств из $(n-1)$ элементов не удовлетворяет (10). Эти ребра образуют дерево, следовательно, у $G(M)$ имеется по крайней мере $(m+1)$ узлов. Следующая матрица $(n+1) \times n$ реализует $s(F_n^n)$:

$$\begin{matrix} 0 & 0 \ldots 0 \\ 1 & 0 \ldots 0 \\ 0 & 1 \ldots 0 \\ \vdots & \\ 0 & 0 \ldots 1 \end{matrix}$$

Все вышеприведенное доказывает следующую теорему.

Теорема 5 [5]:

$$s(F_1^n) = 2, \quad s(F_2^n) = \left[\frac{1 + \sqrt{1 + 8n}}{2}\right],$$

$$s(F_n^n) = n + 1, \quad s(F_{n-1}^n) = n.$$

Из леммы 1 следует и $s(F_3^n) \geqslant n$.

Предположение. Если $n \geqslant 7$, то $s(F_3^n) = n$. Для некоторых случаев это предположение уже доказано.

Теорема 6. (Фюреди). Если $n = h^r$, то $s(F_3^n) = n$.

Доказательство. Построим матрицу $h^r \times h^r$, которая реализует F_3^n. Рассмотрим $2r$-мерное множество h^r-векторов V, состоящее из нулей и единиц. Покажем методом полной индукции, что это множество (без вектора, состоящего только из нулей) можно разделить на трехэлементные подмножества $A_1, \ldots, A_{(h^r-1)/3}$ так, что для трех векторов v_1, v_2, v_3, попадающих в одно подмножество, выполняется $v_1 + v_2 = v_3$ относительно координатного сложения по модулю 2. В этом случае уже следует, что $v_1 + v_3 = v_2$, $v_2 + v_3 = v_1$. Для $r = 1$ утверждение верно: $(0, 1) + (1,0) = (1,1)$. Предположим, что для r это также верно, и докажем утверждение для $r + 1$.

Из всех множеств $A = \{v_1, v_2, v_3\}$, записанных для r_3, сделаем четыре новых:

1) к каждому окончанию векторов v_1, v_2, v_3 припишем 00;

2) после v_i, v_{i+1}, v_{i+2} припишем 01, 10 и 11 при $i = 1, 2, 3$ ($v_4 = v_1$, $v_5 = v_2$). Из тождественно нулевого вектора длиной $2r$ сделаем трехэлементное множество, приписав к нему 01, 10 и 11.

Легко можно убедиться, что таким образом все $2(r+1)$-мерные векторы из 0, 1 получаются только один раз. Этим и заканчивается желаемое разбиение.

Следовательно, доказано, что существует следующее разбиение:

$$V = A_0 \cup A_1 \cup \dots \cup A_{(h^r-1)/3}, A_i \cap A_j = \varnothing,$$

$$A_0 = \{(0, \dots, 0)\} \quad (0 \leqslant i < j \leqslant (h^r - 1)/3), \quad (11)$$

$$|A_i| = 3 \quad (1 \leqslant i \leqslant (h^r - 1)/3),$$

а сумма любых двух элементов A_i дает третье условие:

$$(1 \leqslant i \leqslant (h^r - 1)/3). \quad (12)$$

Пусть $v \in V$, а $A + v$ — множество векторов вида $a + v$ (покоординатно по модулю 2), где $a \in A$. Строки и столбцы матрицы M пронумеруем элементами V (нумеруем целыми числами, у которых рассмотренные векторы являются представлениями в двоичной системе счисления). Пусть v_2-й элемент v_1-й строки M будет числом i, определенным отношением

$$v_1 + v_2 \in A_i. \quad (13)$$

Такое i существует на основании (11) и определяется однозначно. Таким образом, v_1-й элемент v_1-й строки будет равен нулю (v_1 — нулевой вектор).

Относительно M необходимо показать, что в этой матрице никаких два столбца не являются ключами, иначе говоря, к произвольно выбранным столбцам v_2 и $v_2 \neq v_2$ всегда можно найти строки v_1 и $v_1' \neq v_1$, которые совпадают в столбцах v_2 и v_2'. В силу (11) для любого целого $i \neq 0$ выполняется $v_2 + v_2' \in A_i$. Пусть a — еще какой-то элемент A_i и пусть также $v_1 = a + v_2$ и $v_1' = a + v_2'$. Отсюда

$$v_1 + v_2 = a, \quad v_1' + v_2' = a, \quad v_1 + v_2' = a + v_2 + v_2',$$

$$v_1' + v_2 = a + v_2 + v_2'.$$

Поскольку a и $a + v_2 + v_2' \in A_i$, то (по определению (13)) на пересечении строк v_1, v_1' и столбцов v_2, v_2' матрицы M везде стоит i. Действительно, столбцы v_2 и v_2' ключа не образуют.

Теперь возьмем три разных столбца—v_1, v_2' и v_2''. Предположим от противного, что какие-то строки v_1 и $v_1' \neq v_1$ в этих столбцах совпадают, т. е., используя определение (13), примем, что

$$v_1 + v_2 \in A_i, \quad v_1' + v_2 \in A_i, \quad (14)$$

$$v_1 + v_2' \in A_j, \quad v_1' + v_2 \in A_j, \quad (15)$$

$$v_1 + v_2'' \in A_k, \quad v_1' + v_2'' \in A_k \quad (16)$$

выполняются для каких-то i, j ($k \neq 0$). В силу (14) следует, что $v_1 + v_2 + v_1' + v_2 = v_1 + v_1' \in A_i$. Аналогично из (15) и (16) следует, что $v_1 + v_1' \in A_j$ и $v_1 + v_1' \in A_k$. Иначе говоря, $A_i = A_j = A_k$, т. е. $i = j = k$. Так как v_2, v_2' и v_2'' разные, поэтому $a = v_1 + v_2$, $b = v_1 + v_2'$ и $c = v_1 + v_2''$ будут разными элементами A_i. Отсюда следует, что $v_1' + v_2$, $v_1' + v_2'$ и $v_1' + v_2''$ — также разные. Если $v_1' + v_2 \neq \neq v_1 + v_2$, то либо 1) $v_1' + v_2 = b$, либо 2) $v_1' + v_2 = c$.

В случае 1) $v_1' + v_2'$ не может быть a, поскольку иначе $v_1' + v_2''$ может быть только c, что приведет к противоречию с $v_1 + v_2'' = v_1' + v_1''$, поэтому $v_1' + v_2'' = c$. Значит, в этом случае $v_1' + v_2 = b$, $v_1' + v_2' = c$, $v_1' + v_2'' = a$. Окончательное противоречие следует из таких ревенств:

$$c = a + b = (v_1 + v_2) + (v_1 + v_2) = v_1 + v_1',$$

$$a = b + c = (v_1 + v_2') + (v_1' + v_2') = v_1 + v_1'.$$

В случае 2) также можно получить противоречие, подобное вышеприведенному. Таким образом, невозможно, чтобы исходное множество выполнялось, т. е. три столбца не образуют ключа.

Теорема 6 доказана.

Фюреди удалось доказать предположение и для числа вида $n = 12k + 1$ и $n = 12k + h$ с помощью результатов Ганани [6].

3. О КОЛИЧЕСТВЕ ЗАВИСИМОСТЕЙ

Введем понятие зависимости одного множества столбцов от другого. Эвристически можно сказать, что B зависит от A (A и B — множества столбцов матрицы M), если элементы любой строки M, попадающие в A, однозначно определяют элементы, попадающие в B. Точнее, пару (A, B) из множества столбцов M назовем зависимостью, если любые две строки матрицы M, которые совпадают на местах, соответствующих A, совпадают и на местах, соответствующих B. Обозначение $A \to B$ будем использовать для указания того, что (A, B) является зависимостью. Из определений видно, что условие $A \to$ (все столбцы M) означает, что A является ключом.

Здесь, как и при описании ключей, будем обращать внимание только на некоторые определяющие, или существенные, зависимости. Легко можно показать, что

$$A \to A, \quad (17)$$

$$\text{если } A \to B \text{ и } A \subset A', \text{ то } A' \to B, \quad (18)$$

$$\text{если } A \to B \text{ и } B' \subset B, \text{ то } A \to B'. \quad (19)$$

Если не обращать внимания на излишние зависимости, то на основе (17)—(19) можно выделить существенную зависимость $A \to B$, если

$$A \neq B, \quad (20)$$

нет такого $A' \subset A$, $A' \neq A$, что $A' \to B$, (21)

нет такого $B' \supset B$, $B' \neq B$, что $A \to B'$. (22)

Из существенных зависимостей на основе (17)—(19) можно вывести все остальные зависимости, поэтому число существенных зависимостей можно рассматривать как меру сложности матрицы. В дальнейшем рассмотрим максимум сложности матриц, обладающих n столбцами. Пусть $N(n)$ — максимальное число существенных зависимостей матриц из n столбцов.

Легко построить такую матрицу, в которой все существенные зависимости имеют вид $A \to A \cup \{x\}$, где x — последний столбец, $x \notin A$. Число существенных зависимостей в ней будет 2^{n-1}.

Теорема 7 [7]. Если взять логарифм по основанию 2, то

$$2^n\left(1 - \frac{h}{\log e}\cdot\frac{\log\log n}{\log n}(1+\sigma(1)\right) \leqslant$$

$$\leqslant N(n) \leqslant 2^n\left(1 - \frac{1}{n+1}\right). \qquad (23)$$

Доказательство. Нижняя оценка. Пусть $q_1, \ldots, q_k$ — положительные целые, $\sum_{i=1}^{k} q_i = n$. Матрицу $Q_i\ (\binom{q_i}{2} + 1) \times q_i$ определим следующим образом. Первая строка Q_i состоит из единиц, остальные строки содержат $(q_i - 2)$ единиц. На оставшиеся два места i-й строки впишем i $(2 \leqslant i \leqslant \binom{q_i}{2} + 1)$. Строки матрицы $M = Q_1 \times \ldots \times Q_k$ образуем, написав всеми возможными способами строки матрицы Q_i друг за другом. Иначе говоря, одну строку M получим таким образом: после произвольной строки Q_1 запишем произвольную строку Q_2, а далее запишем произвольную строку Q_3 и т. д. Тогда число столбцов M будет $\sum_{i=1}^{k} qi = n$.

В качестве примера рассмотрим следующую конструкцию (при $n = 5$, $k = 2$, $q_1 = 3$, $q_2 = 2$):

1	1	1	:	1	1
1	1	1	:	2	2
1	2	2	:	1	1
1	2	2	:	2	2
3	1	3	:	1	1
3	1	3	:	2	2
4	4	1	:	1	1
4	4	1	:	2	2

Пусть Q_i^* — индекс тех столбцов в M, которые соответствуют Q_i, т. е. $Q_i^* = \{q_1 + \ldots q_{i-1}, \ldots, q_1 + + \ldots q_i\}$. Можно легко проверить, что существенными зависимостями $A \to B$ являются те, для которых

$$\min_{B = A \cup (\cup Q_i)} (q_i - |A \cap Q_i^*|) = A,$$

где объединение берется по тем i, для которых выполняется $|A \cap Q_i^*| = q_i - 1$. Число таких пар будет

$$\prod_{i=1}^{k} (2^{q_i} - 1) - \prod_{i=1}^{k} (2^{q_i} - q_i - 1). \qquad (24)$$

Это дает нижнюю оценку для $N(n)$. Для вышеприведенного примера $N(5) \geqslant 17$, что лучше, чем очевидная оценка $2^h = 16$.

Для больших n выберем значение

$$q = q(n) = \left\lfloor \log n - \log\left(\frac{1}{\log e}(\log\log n - \right.\right.$$

$$\left.\left. - \log\log\log n - \log\log e - 1)\right)\right\rfloor,$$

а $k(n)$ и $r(n)$ определим делением с остатком $n = q(n)k(n) + r(n)$, $0 \leqslant r(n) < q(n)$.

Наконец, пусть

$$q_1 = q_2 = \ldots = q_r = q + 1, \quad q_{r+1} = \ldots = q_k = q.$$

Отсюда и из (24) после достаточно сложных вычислений (см. [7]) получим левую часть (23).

Верхняя оценка. Пусть F — семейство таких множеств A, B, для которых $A \to B$ — существенная зависимость. Тогда $A \subset B$, но $A \neq B$ (на основе (20)—(22)). Возьмем C из условия $A \subset C \subset B$, $|C| = |A| + 1$ — множество. Можно легко убедиться в том, что $C \in F$, поскольку в противном случае $C \to D$ было бы существенной зависимостью. Таким образом, получаем противоречие.

Такое C можно получить не более чем из n членов F, поэтому существует такое множество $|F|/n$, которое не является элементом F. Отсюда следует неравенство

$$|F| + \frac{|F|}{n} \leqslant 2^n,$$

что эквивалентно правой части (23).

Теорема 7 доказана.

В правой части (23) член $\frac{1}{n+1}$ по порядку можно улучшить на $\frac{2}{n}$, так как большинство элементов F имеет приблизительно $\frac{n}{2}$ элементов, а следовательно, и число элементов соответствующего множества C примерно такое же.

Получение верхней оценки можно еще улучшить за счет большой разницы между вторыми членами в (23).

Окончание см. на стр. 69.

Окончание. Начало см. на стр. 37.

СПИСОК ЛИТЕРАТУРЫ

1. Codd E. F. A relational model of data for large shared data banks.— Comm. ACM, 1970, N 13, p. 377—387.
2. Demetrovics J. On the equivalence of candidate Keys with Sperner sytems.— Acta Cybernet., 1979, N 4, p. 247—252.
3. Sperner E. Ein Satz über Untermengen einer endlichen Menge.— Math. Z., 1928, N 27, p. 544—548.
4. Demetrovics J., Gyepesi Gy. On the functional dependency andsome generalization of it. — Acta Cybernet., 1981, N 5, p. 296—305.
5. Demetrovics J., Catona G. O. H. Extremal combinatorial problems in relational data base. Fundamentals of Computation Theory. — Lect. Notes Comput. Sci., 1975, N 117, p. 110—119.
6. Hanani H. The existence and construction of balanced incomplete block designs.— Ann. Math. Statist., 1961, N 32, p. 361—386.
7. On the number of maximal dependencies in a data base relation of fixed order / J. Békéssy, J. Demetrovics, L. Hannák, P. Franke, G. O. H. Katona.— Discrete Math., 1980, N 30, p. 83—88.

Поступила 20.01.83