О дуализации над произведением частичных порядков*

Е. В. Дюкова[1], Г. О. Масляков[2], П. А. Прокофьев[3]
edjukova@mail.ru; gleb-mas@mail.ru; p_prok@mail.ru
[1]ФИЦ «Информатика и управление» РАН, Россия, г. Москва, ул. Вавилова, 44/2
[2]МГУ им. М. В. Ломоносова, Россия, г. Москва, Ленинские горы, 1
[3]ИМАШ РАН им. А. А. Благонравова, Россия, г. Москва, Малый Харитоньевский пер., 4

Рассматривается одна из центральных труднорешаемых задач логического анализа данных — дуализация над произведением частичных порядков. Исследуется важный частный случай, когда каждый порядок является цепью. Показано, что поставленная задача сводится к поиску неприводимых покрытий булевой матрицы (дуализации булевой матрицы), специальным образом построенной по исходным данным. Приведены результаты численных экспериментов, базирующиеся на эффективном «в типичном случае» асимптотически оптимальном поиске неприводимых покрытий булевой матрицы. Ранее для решения рассматриваемой задачи предлагался подход, представляющий интерес, в основном, для теории и имеющий целью построение инкрементальных алгоритмов с квазиполиномиальными временными оценками «для худшего случая».

Ключевые слова: *дуализация; произведение частичных порядков; цепь; покрытие булевой матрицы; ассимптотически оптимальный алгоритм*

DOI: 10.21469/22233792.3.4.02

1 Введение

Логический анализ данных основан на решении сложных в вычислительном плане задач, что естественно обусловлено применением дискретного аппарата. Как правило, возникают задачи, которые в теории алгоритмической сложности дискретных задач называют труднорешаемыми. Особой сложностью отличаются перечислительные задачи, в которых требуется найти (перечислить) все решения, при этом число решений растет экспоненциально с ростом размера задачи (размера входа). Одной из главных перечислительных задач считается дуализация над произведением частичных порядков. Ниже приведена ее формулировка.

Пусть $P = P_1 \times \dots \times P_n$, где $P_1, \dots, P_n$ — конечные частично упорядоченные множества. Считается, что элемент $x = (x_1, \dots, x_n) \in P$ следует за элементом $y = (y_1, \dots, y_n) \in P$, если x_i следует за y_i при $i = 1, 2, \dots, n$. Для обозначения того, что $y \in P$ следует за $x \in P$ и $x \neq y$, далее используется запись $x \prec y$.

Пусть $R \subseteq P$, $R^+ = R \cup \{x \in P | \exists a \in P,\ a \prec x\}$. Задача построения двойственного к R множества $I(R)$, состоящего из элементов $a \in P \setminus R^+$ таких, что для любого $x \in P \setminus R^+$ $+$, $x \neq a$, отношение $a \prec x$ не выполняется, называется дуализацией над произведением частичных порядков. Элемент из $I(R)$ называется *максимальным независимым* от R элементом множества P.

Одним из наиболее востребованных является случай, когда каждое P_i является цепью, т. е. любые два элемента в P_i сравнимы. Если $P_i = \{0, 1\}$ при $i \in \{1, 2, \dots, n\}$ и $0 \prec 1$, то рассматриваемая задача — это перечисление максимальных независимых подмножеств

*Работа выполнена при финансовой поддержке РФФИ, проект №16-01-00445.

вершин гиперграфа с n вершинами и $|R|$ ребрами (дуализация гиперграфа). Эквивалентными задачами являются следующие две задачи. Во-первых, это построение сокращенной дизъюнктивной нормальной формы монотонной булевой функции от n переменных, заданной конъюнктивной нормальной формой из $|R|$ элементарных дизъюнкций (дуализация монотонной булевой функции). Во-вторых, это поиск неприводимых покрытий булевой матрицы из $|R|$ строк и n столбцов (дуализация булевой матрицы). Если R состоит из попарно несравнимых элементов, то дуализация монотонной булевой функции — это построение множества «нижних» единиц этой функции при условии, что задано множество ее «верхних» нулей.

Важность дуализации обусловлена большим числом приложений, среди которых следует выделить логический анализ данных в распознавании (машинное обучение по прецедентам), поиск ассоциативных правил в базах данных (data mining), решение монотонных систем неравенств (целочисленное и стохастическое программирование), пересечение матроидов (комбинаторная оптимизация и символьный анализ электронных цепей), соединение вершин наименьшим набором графов (теория надежности), покрытие линейного пространства подпространствами (криптография), поиск эффективных точек дискретных вероятностных распределений (стохастическое программирование), упаковка точек (data mining), поиск минимальных тестов (теория управляющих систем).

Теоретические оценки эффективности алгоритмов дуализации базируются на оценке сложности одного шага [1]. Наиболее эффективным считается алгоритм, который имеет полиномиальный от размера входа шаг. Хотя задача поставлена еще в 1960-х гг. [2], полиномиальные алгоритмы удалось построить лишь для некоторых частных случаев дуализации, поэтому требования к алгоритму были ослаблены. Обозначились два направления исследований.

Первое направление, разрабатываемое в основном за рубежом, основано на построении так называемых инкрементальных алгоритмов, когда алгоритму разрешено просматривать решения, найденные на предыдущих шагах. При этом оценки сложности шага алгоритма даются для худшего случая (самого сложного варианта задачи). В [3] построен алгоритм дуализации монотонной булевой функции с квазиполиномиальным шагом, определяемым не только размером входа задачи, но и размером ее выхода. Такой алгоритм интересен исключительно для теории, поскольку в худшем случае число решений дуализации (размер выхода задачи) растет экспоненциально с ростом размера ее входа.

Второе направление исследований основано на построении асимптотически оптимальных алгоритмов дуализации булевой матрицы (предложено в [4]). В этом случае алгоритму разрешено делать лишние полиномиальные шаги при условии, что их число должно быть достаточно мало по сравнению с числом всех решений задачи (числом неприводимых покрытий булевой матрицы). В результате удалось построить алгоритмы дуализации булевой матрицы, эффективные в типичном случае (для почти всех вариантов задачи). Эти алгоритмы имеют теоретическое обоснование и показывают хорошие результаты на практике. В дальнейшем подход был применен для построения асимптотически оптимальных алгоритмов дуализации монотонной булевой функции, а также для более общих задач преобразования нормальных форм двузначной функции многозначной логики [5–9].

В настоящее время отечественные асимптотически оптимальные алгоритмы поиска неприводимых покрытий булевой матрицы являются мировыми лидерами по скорости счета [9]. Эти алгоритмы позволяют решать задачи значительных размеров, что подтверждают эксперименты на большом количестве разнотипных данных. Данные для тестирования предоставлены японскими учеными Murakami и Uno [10]. Следует отметить, что алгоритмы дуализации гиперграфа, предложенные в [10], являются в силу эквивалентности задач асимптотически оптимальными алгоритмами поиска неприводимых покрытий булевой матрицы. Однако эти алгоритмы, а также другие известные алгоритмы дуализации гиперграфа, имеющие иные конструктивные особенности, уступают по скорости счета последним отечественным разработкам, представленным в [9].

В настоящей работе рассматривается случай, когда $P_i = \{0, 1, \ldots, k-1\}$, $k \geqslant 2$, $i = 1, 2, \ldots, n$, и элементы в P_i упорядочены в порядке возрастания, т.е. $0 \prec 1 \prec 2 \prec \cdots \prec k-1$.

В разд. 2 введено понятие упорядоченного тупиковым σ-покрытием целочисленной матрицы, являющееся обобщением понятия неприводимого покрытия булевой матрицы, и показано, что исходная задача сводится к перечислению покрытий общего вида для матрицы $L(R^+)$, строками которой являются наборы из R^+. В разд. 3 приведено утверждение, согласно которому задача построения упорядоченных тупиковых σ-покрытий матрицы $L(R^+)$ может быть решена на основе построения специальных неприводимых покрытий булевой матрицы из $|R|$ строк и kn столбцов. Для поиска искомых неприводимых покрытий в работе построен алгоритм RUNC-M+, являющийся модификацией асимптотически оптимального алгоритма поиска неприводимых покрытий булевой матрицы RUNC-M из [9]. Его описание на псевдокоде, а также описание общей схемы работы асимптотически оптимального алгоритма дуализации булевой матрицы приведены в разд. 4 и 5. Результаты тестирования алгоритма RUNC-M+ рассмотрены в разд. 6.

Ранее в работе [11] для случая, когда каждое P_i является цепью и $|P_i| \geqslant 2$, на базе алгоритма, предложенного в [3], построен квазиполиномиальный инкрементальный алгоритм. Однако результаты тестирования этого алгоритма в [11] не приведены.

2 Основные понятия

Пусть L — матрица с n столбцами и элементами из $\{0, 1, \ldots, k-1\}$, $k \geqslant 2$, и пусть $\sigma = (\sigma_1, \ldots, \sigma_r)$, $r \leqslant n$, $\sigma_i \in \{0, 1, \ldots, k-2\}, k \geqslant 2$.

Упорядоченным тупиковым σ-покрытием матрицы L называется набор H из r различных столбцов этой матрицы такой, что подматрица L^H матрицы L, образованная столбцами набора H, обладает следующими двумя свойствами: (1) L^H не содержит строку $(\sigma_1, \ldots, \sigma_r)$; (2) если $t \in \{1, 2, \ldots, r\}$, то L^H содержит хотя бы одну строку $(\beta_1, \ldots, \beta_r)$ такую, что β_t непосредственно следует за σ_t и $\beta_j = \sigma_j$ при $j \in \{1, \ldots, r\} \setminus \{t\}$.

Упорядоченное тупиковое $(0, 0, \ldots, 0)$-покрытие H булевой матрицы L называется неприводимым покрытием. Если L — булева матрица, $\sigma = (0, \ldots, 0)$ и выполнено условие (1), то H — покрытие матрицы L (H покрывает строки матрицы L). Если L — булева матрица, $\sigma = (0, \ldots, 0)$ и выполнено условие (2), то H — совместимый набор столбцов матрицы L. Если же условие (2) не выполнено, то H — несовместимый набор столбцов матрицы L. Совместимый набор столбцов булевой матрицы называется максимальным, если он не содержится ни в каком другом совместимом наборе столбцов этой матрицы.

Пусть H — совместимый набор столбцов булевой матрицы L. Будем говорить, что столбец h матрицы L совместим с H, если $H \cup \{h\}$ — совместимый набор столбцов матрицы L.

Подматрица булевой матрицы называется единичной, если в каждой строке и в каждом столбце этой матрицы в точности один элемент равен 1. Единичная подматрица булевой матрицы L называется максимальной, если она не содержится ни в какой другой единичной подматрице матрицы L. Из условия, что H — максимальный совместимый набор столбцов матрицы L длины r, следует, что L^H содержит хотя бы одну максимальную единичную подматрицу порядка r.

Будем говорить, что строка $(a_1, a_2, \ldots, a_n)$ булевой матрицы L охватывает строку $(b_1, b_2, \ldots, b_n)$ этой матрицы, если $a_j \geqslant b_j$ при $j = 1, 2, \ldots, n$. Заметим, что множество $(0, 0, \ldots, 0)$-покрытий булевой матрицы не меняется при выбрасывании из нее охватывающих строк.

Пусть $P = P_1 \times \cdots \times P_n,\ P_i = \{0, 1, \ldots, k-1\},\ k \geqslant 2,\ R \subseteq P$. Обозначим через L_R матрицу, строками которой являются элементы множества R, через L_{R^+} матрицу, строками которой являются элементы множества R^+.

Пусть $\sigma = (\sigma_1, \ldots, \sigma_n)$ — набор из P, в котором элемент с номером t, $t \in \{j_1, \ldots, j_r\}$, не является максимальным в P_t, а элемент с номером t, $t \notin \{j_1, \ldots, j_r\}$, является максимальным в P_t. Очевидным является

Утверждение 1. *Набор σ является максимальным независимым от R набором тогда и только тогда, когда набор столбцов матрицы L_{R^+} с номерами $j_1, \ldots, j_r$ является упорядоченным тупиковым $(\sigma_{j_1}, \ldots, \sigma_{j_r})$-покрытием.*

При $k = 2$ матрица L_R получается из матрицы L_{R^+} удалением охватывающих строк, поэтому из утверждения 1 сразу следует

Утверждение 2. *Если $k = 2$, то набор σ является максимальным независимым от R набором тогда и только тогда, когда набор столбцов матрицы L_R с номерами $j_1, \ldots, j_r$ является неприводимым покрытием.*

3 Сведение задачи построения упорядоченных тупиковых σ-покрытий целочисленной матрицы к задаче построения неприводимых покрытий булевой матрицы

Пусть $L = (a_{ij}),\ i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n,$ — матрица с элементами из $\{0, 1, \ldots, k-1\},\ k \geqslant 2$. Пусть $a \in \{0, 1, \ldots, k-1\}$. Положим

$$\delta(a_{ij}, a) = \begin{cases} 1, & \text{если } a_{ij} > a\,; \\ 0, & \text{если } a_{ij} \leqslant 0\,, \end{cases}$$

для $i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n$.

Построим булеву матрицу L^*, состоящую из m строк и $k \times n$ столбцов, в которой строка с номером i, $i = 1, 2, \ldots, m$, имеет вид:

$$(\delta(a_{i1}, 0), \ldots, \delta(a_{i1}, k-1), \delta(a_{i2}, 0), \ldots, \delta(a_{i2}, k-1), \ldots, \delta(a_{in}, 0), \ldots \delta(a_{in}, k-1)).$$

Нетрудно заметить, что столбцу с номером j, $j \in \{0, 1, \ldots, n\}$, исходной матрицы L соответствует группа из k столбцов матрицы L^* с номерами $k(j-1)+1,\ k(j-1)+2, \ldots,\ kj$. Такие столбцы назовем родственными. Через $P(L^*)$ обозначим множество всех неприводимых покрытий матрицы L^*. Несложно доказывается

Утверждение 3. *Пусть $\sigma = (\sigma_1, \ldots, \sigma_r)$, $r \leqslant n$, $\sigma_i \in \{0, 1, \ldots, k-2\}$, при $i = 1, 2, \ldots, r$, $k \geqslant 2$. Набор из r различных столбцов матрицы L с номерами $j_1,\ j_2, \ldots, j_r$ является упорядоченным тупиковым σ-покрытием тогда и только тогда, когда выполнены два следующих условия:* (1) *набор столбцов матрицы L^* с номерами $t_1, t_2, \ldots, t_r$, где $t_i = (j_i - 1)k + \sigma_i + 1$, принадлежит $P(L^*)$;* (2) *если $i \in \{1, 2, \ldots, r\}$ и $t_i < kj_i$, то в $P(L^*)$ нет набора столбцов с номерами $t_1, \ldots, t_{i-1}, q_i, t_{i+1}, \ldots, t_r$, где $q_i \in [t_i + 1, kj_i]$.*

Из утверждения 3 следует, что задача построения множества $P(L)$ сводится к построению подмножества $\widetilde{P}(L^*)$ множества $P(L)$, состоящего из всех таких неприводимых покрытий, которые, во-первых, не содержат столбцов с родственными номерами и, во-вторых, удовлетворяют некоторому дополнительному условию (2), которое назовем условием старшинства. Элементы множества $\widetilde{P}(L^*)$ назовем правильными неприводимыми покрытиями матрицы L^*.

4 Асимптотически оптимальный алгоритм поиска неприводимых покрытий булевой матрицы

Асимптотически оптимальный алгоритм A построения множества всех неприводимых покрытий $P(L)$ булевой матрицы L размера $m \times n$ действует по следующей схеме. На каждом шаге строится максимальный совместимый набор столбцов H матрицы L. Некоторые максимальные совместимые наборы столбцов могут строиться неоднократно. Проверка на повторяемость осуществляется путем просмотра строк подматрицы L^H матрицы L, образованной столбцами набора H. При этом на каждом шаге выполняется не более чем d элементарных операций, где d ограничено сверху полиномом от m и n. Под элементарной операцией понимается просмотр одного элемента матрицы. Основное требование: число шагов $N_A(L)$ алгоритма A должно быть асимптотически равно $|P(L)|$ при $n \to \infty$ для почти всех булевых матриц L размера $m \times n$. Указанное требование означает следующее. Существуют две положительные бесконечно убывающие функции $\alpha(n)$ и $\beta(n)$ такие, что для всех достаточно больших n имеет место

$$1 - \frac{|M|}{|M_{mn}|} \leqslant \alpha(n),$$

где M_{mn} — множество всех булевых матриц размера $m \times n$; M — множество матриц L из M_{mn}, для которых выполнено

$$1 - \beta(n) \leqslant \frac{|P(L)|}{N_A(L)} \leqslant 1 + \beta(n).$$

Таким образом, асимптотически оптимальный алгоритм дуализации булевой матрицы имеет «лишние» полиномиальные шаги. Шаг считается лишним в двух случаях: (1) построенный максимальный совместимый набор столбцов H уже строился на предыдущих шагах; (2) набор столбцов H ранее не строился, но он не является покрытием. Для почти всех булевых матриц размера $m \times n$ число лишних шагов должно иметь более низкий порядок роста по сравнению с числом неприводимых покрытий при росте размера задачи.

Рассматриваемые алгоритмы поиска неприводимых покрытий булевой матрицы L построены для случая, когда число строк матрицы L существенно меньше числа ее столбцов. Обоснование подхода опирается на технику получения асимптотических оценок числа неприводимых покрытий, которая первоначально была предложена в работах В. А. Слепян и В. Н. Носкова, а затем развита в работах Дюковой [4–8] и Андреева [12]. В частности, было показано, что при $\log m \leqslant (1 - \varepsilon) \log n, 0 < \varepsilon < 1, n \to \infty$, для почти всех булевых матриц L размера $m \times n$ число единичных подматриц матрицы L асимптотически равно мощности $P(L)$.

Для наглядности работу асимптотически оптимального алгоритма дуализации булевой матрицы L можно представить в виде обхода дерева решений (ДР) в глубину: корнем ДР является пустой набор; вершины — совместимые наборы столбцов; висячие вершины —

максимальные совместимые наборы столбцов, которые либо являются впервые найденными неприводимыми покрытиями, либо соответствуют лишним шагам.

Первоначально матрица L рассматривается в качестве текущей матрицы. Построение ДР начинается с вершины, которая порождается некоторым ненулевым столбцом матрицы L. При построении каждой из остальных вершин в роли текущей матрицы выступает подматрица матрицы L. Для построения дочерней вершины внутренней вершины H алгоритм по определенному правилу выбирает в текущей матрице ненулевой столбец и строит новую вершину $H \cup h$, где h — соответствующий столбец матрицы L. После построения вершины $H \cup h$ текущая матрица изменяется. Из нее, как правило, удаляются все строки, покрытые столбцом h (имеющие единицу в пересечении с этим столбцом), и удаляются столбцы, порожденные теми столбцами матрицы L, которые несовместимы с набором $H \cup h$. Дополнительно могут удаляться и другие строки и столбцы. Если новая текущая матрица не содержит единичных элементов, то вершина $H \cup h$ становится висячей. Если набор $H \cup h$ найден впервые и является покрытием, то множество неприводимых покрытий, построенных на предыдущих шагах, пополняется; в противном случае это множество не меняется и шаг объявляется лишним. В любом случае далее либо происходит переход к новому шагу путем возврата на более высокий уровень ДР, либо алгоритм заканчивает работу. Если же новая текущая матрица содержит единичные элементы, то $H \cup h$ — внутренняя вершина ДР и процесс построения ветви дерева продолжается.

Время работы асимптотически оптимального алгоритма в большой степени зависит от сложности ДР (числа вершин в ДР), которое строит этот алгоритм.

На данный момент наилучшие результаты по скорости счета показывает алгоритм RUNC-M. Этот алгоритм строит каждый максимальный совместимый набор только один раз. В RUNC-M применяется «жадная» стратегия выбора столбцов для построения дочерних вершин в ДР. Всякий раз после построения очередной внутренней вершины H в текущей матрице выбирается строка i, имеющая наименьшее число единиц. Дочерняя для H вершина строится путем добавления к набору H столбца матрицы L, имеющего наименьший номер среди столбцов, образующих текущую матрицу и покрывающих строку i. Сложность шага алгоритма $O(qmn), q = \min(m, n)$.

5 Алгоритм дуализации произведения цепей RUNC-M+

Согласно утверждению 1 поиск максимальных независимых от R наборов сводится к поиску упорядоченных тупиковых σ-покрытий матрицы L_{R^+}.

Преобразуем матрицы L_R и L_{R^+} соответственно в булевы матрицы L^*_R и $L^*_{R^+}$ способом, описанным в разд. 2. Согласно утверждению 3, поиск упорядоченных тупиковых σ-покрытий матрицы L_{R^+} сводится к поиску правильных неприводимых покрытий матрицы $L^*_{R^+}$. Нетрудно видеть, что каждая строка из $L^*_{R^+}$, не содержащаяся в L^*_R, охватывает хотя бы одну строку из L^*_R. Следовательно, набор столбцов с номерами $j_1, \ldots, j_r$ матрицы $L^*_{R^+}$ является неприводимым покрытием тогда и только тогда, когда набор столбцов с номерами $j_1, \ldots, j_r$ матрицы L^*_R является неприводимым покрытием.

Построенный в настоящей работе алгоритм поиска правильных неприводимых покрытий матрицы $L^*_{R^+}$ является модификацией алгоритма дуализации булевой матрицы RUNC-M и назван RUNC-M+.

Алгоритм RUNC-M+ описывается ниже рекурсивной процедурой RUNCM $(L^*_R; H; D; C)$, первый вызов которой осуществляется с параметрами $H = \varnothing$, $D = \{1, 2, \ldots, m\}$, $C = \{1, 2, \ldots, kn\}$.

Процедура RUNCM ($L_R^*; H; D; C$)

1: $C^{\min} := \{j \in C | a_{ij} = 1\}$, где i — номер строки из D с минимальным числом элементов, равных 1
2: **для всех** $j \in C^{\min}$
3: $\quad C := C \setminus \{j\}$
4: $\quad H := H \cup \{j\}$
5: $\quad$ Исключить из D номера строк, покрытых столбцом с номером j
6: $\quad$ **если** $D = \varnothing$ и набор столбцов с номерами из H удовлетворяет условию старшинства, **то**
7: $\quad\quad$ Сохранить в $P(L_R^*)$ набор столбцов с номерами из H (найден новый элемент из $P(L_R^*)$)
8: $\quad$ **иначе**
9: $\quad\quad$ **если** $D \neq \varnothing$ **то**
10: $\quad\quad\quad$ Исключить из C номера столбцов, не совместимых со столбцами с номерами из H
11: $\quad\quad\quad$ **если** $C \neq \varnothing$, **то**
12: $\quad\quad\quad\quad$ Вызвать RUNCM(L_R^*, H, D, C)
13: $\quad$ Отменить изменения, внесенные на шагах 4 и 5.

6 Эксперименты

Были проведены эксперименты с целью выяснить, как меняется время работы алгоритма RUNC-M+ в случае с определенными конфигурациями матриц, а именно: с «вытянутыми» по горизонтали, с «вытянутыми» по вертикали и с квадратными. Также исследовалось влияние размерности частичного порядка на время счета. Для экспериментов генерировались случайные матрицы, причем каждый элемент матрицы генерировался независимо из равномерного дискретного распределения. Среднее время работы алгоритма бралось по 20 независимым матрицам.

В табл. 1 приведены результаты работы алгоритма в зависимости от размера целочисленной матрицы L_R и размерности частичного порядка.

При исследовании времени работы алгоритма на «вытянутых» по вертикали матрицах было замечено, что на определенном этапе увеличение числа строк приводит к резкому увеличению скорости работы алгоритма. По-видимому, это происходит из-за существенного уменьшения числа правильных неприводимых покрытий матрицы L_R^*. Примеры данного эффекта приведены в табл. 2 (рассмотрен случай, когда значность k матрицы L_R равна трем).

Из результатов эксперимента видно, что время работы алгоритма RUNC-M+ в большей степени зависит от числа столбцов, чем от числа строк. При увеличении числа столбцов наблюдался экспоненциальный рост времени работы. Кроме того, достаточное увеличение числа строк по отношению к числу столбцов приводит к снижению времени работы алгоритма.

7 Заключение

В статье рассматривается одна из центральных труднорешаемых задач дискретной математики — дуализация над произведением цепей $P_1, \ldots, P_n$, которая, в частности, возникает при конструировании логических процедур классификации по прецедентам. Предполагается, что мощность каждой цепи P_i равна k, $k \geqslant 2$. Поставленная задача имеет

Таблица 1 Результаты работы алгоритма в зависимости от размера целочисленной матрицы L_R и размерности частичного порядка

Размер/значность k матрицы L_R	Размер матрицы L_R^*	Время работы алгоритма RUNC-M+, с
$20 \times 30/3$	20×90	1,784
$20 \times 35/3$	20×105	5,152
$20 \times 40/3$	20×120	14,226
$20 \times 45/3$	20×135	39,862
$10 \times 10/3$	10×30	0,0000033
$15 \times 15/3$	15×45	0,01
$20 \times 20/3$	20×60	0,113
$25 \times 25/3$	25×75	1,014
$30 \times 30/3$	30×90	15,183
$30 \times 20/3$	30×60	0,26
$35 \times 20/3$	35×60	0,01
$40 \times 20/3$	40×60	0,000768
$45 \times 20/3$	45×60	0,000385
$30 \times 20/4$	30×80	2,216
$35 \times 20/4$	35×80	2,985
$40 \times 20/4$	40×80	3,546
$45 \times 20/4$	45×80	6,728
$10 \times 20/2$	10×40	0,005
$10 \times 20/3$	10×60	0,454
$10 \times 20/4$	10×80	0,112
$10 \times 20/5$	10×100	0,446
$10 \times 20/6$	10×120	0,714
$10 \times 20/7$	10×140	1,871
$10 \times 20/8$	10×160	1,164
$10 \times 20/9$	10×180	2,532
$10 \times 20/10$	10×200	5,788
$20 \times 10/2$	20×20	0,012
$20 \times 10/3$	20×30	0,169
$20 \times 10/5$	20×50	0,032
$20 \times 10/7$	20×70	0,138
$20 \times 10/10$	20×100	0,372
$20 \times 10/15$	20×150	2,384
$20 \times 10/20$	20×200	4,0258

в качестве входа k-значные наборы длины k и при $k = 2$ эквивалентна поиску неприводимых покрытий булевой матрицы размера $m \times n$, где m — число входных наборов. Показано, что в общем случае задача сводится к поиску некоторого подмножества множества неприводимых покрытий булевой матрицы размера $m \times kn$. Приведены результаты численных экспериментов, базирующиеся на эффективном «в среднем» асимптотически оптимальном перечислении неприводимых покрытий. Предлагаемые построения очевидным образом переносятся и на случай, когда множества P_i имеют различные мощности. Ранее для решения рассматриваемой задачи использовался подход, в основном разрабатываемый за рубежом и представляющий интерес исключительно для теории. Этот подход имеет целью

Таблица 2 Результаты работы алгоритма на «вытянутых» по вертикали матрицах ($k = 3$)

Размер матрицы L_R	Размер матрицы L_R^*	Время работы алгоритма RUNC-M+, с	Среднее значение $\lvert\widetilde{P}(L_R^*)\rvert$
31×20	31×60	0,61	93 750
32×20	32×60	0,68	108 840
33×20	33×60	0,144	11 160
34×20	34×60	0,033	2 800
35×20	35×60	0,015	1 120
64×25	64×75	26,189	2 937 850
65×25	65×75	2,5	239 950
66×25	66×75	0,922	94 090
67×25	67×75	0,122	12 530
68×25	68×75	0,085	8 730
98×40	98×120	35,013	6 181 120
99×40	99×120	4,626	754 050
100×40	100×120	1,198	190 560
101×40	101×120	0,612	99 240
102×40	102×120	0,019	3 100

построение инкрементальных алгоритмов с квазиполиномиальными временны́ми оценками «в худшем случае».

Литература

[1] *Johnson D. S., Yannakakis M., Papadimitriou C. H.* On general all maximal independent sets // Inform. Process. Lett., 1988. Vol. 27. P. 119–123.

[2] *Чегис И. А., Яблонский С. В.* Логические способы контроля электрических схем // Сб. статей по математической логике и ее приложениям к некоторым вопросам кибернетики: Тр. МИАН СССР, 1958. Т. 51. С. 270–360.

[3] *Fredman L., Khachiyan L.* On the complexity of dualization of monotone disjunctive normal forms // J. Algorithm., 1996. Vol. 21. P. 618–628. doi: 10.1006/jagm.1996.0062.

[4] *Дюкова Е. В.* Об асимптотически оптимальном алгоритме построения тупиковых тестов // Докл. АН СССР, 1977. Т. 233. № 4. С. 527–530.

[5] *Дюкова Е. В.* О сложности реализации некоторых процедур распознавания // Ж. вычисл. матем. матем. физ., 1987. Т. 27. № 1. С. 114–127. doi: 10.1016/0041-5553(87)90121-2.

[6] *Дюкова Е. В., Журавлёв Ю. И.* Дискретный анализ признаковых описаний в задачах распознавания большой размерности // Ж. вычисл. матем. матем. физ., 2000. Т. 40. № 8. С. 1264–1278. doi: 10.1016/0041-5553(87)90121-2.

[7] *Дюкова Е. В.* О сложности реализации дискретных (логических) процедур распознавания // Ж. вычисл. матем. матем. физ., 2004. Т. 44. № 3. С. 551–561.

[8] *Дюкова Е. В., Инякин А. С.* Асимптотически оптимальное построение тупиковых покрытий целочисленной матрицы // Математические вопросы кибернетики. — М.: Наука, 2008. № 17. С. 235–246.

[9] *Дюкова Е. В., Прокофьев П. А.* Об асимптотически оптимальных алгоритмах дуализации // Ж. вычисл. матем. матем. физ., 2015. Т. 55. № 5. С. 895–910. doi: 10.1134/S0965542515050103.

[10] *Murakami K., Uno T.* Efficient algorithms for dualizing large-scale hypergraphs // Discrete Appl. Math., 2014. Vol. 170. P. 83–94. doi: 10.1016/j.dam.2014.01.012.

[11] *Boros E., Elbassioni K., Gurvich V., Khachiyan L., Makino K.* Dual-bounded generating problems: All minimal integer solutions for a monotone system of linear inequalities // SIAM J. Comput., 2002. Vol. 31. No. 5. P. 1624–1643. doi: 10.1137/S0097539701388768.

[12] *Андреев А. Е.* Об асимптотическом поведении числа тупиковых тестов и минимальной длины теста для почти всех таблиц // Проблемы кибернетики, 1984. Вып. 41. С. 117–141.

Поступила в редакцию 30.08.2017

About product over partially ordered sets*

E. V. Djukova[1], G. O. Masliakov[2], and P. A. Prokofjev[3]

edjukova@mail.ru; gleb-mas@mail.ru; p_prok@mail.ru

[1]Federal Research Center "Computer Science and Control" of RAS, 44/2 Vavilova Str., Moscow, Russia

[2]Lomonosov Moscow State University, 1 Leninskie Gory, Moscow, Russia

[3]Blagonravov IMASH of RAS, 4 Malyj Kharitonyevskiy per., Moscow, Russia

One of the central, intractable problems of logical data analysis — the dualization over the product of partial orders — is considered. An important special case is investigated, when each order is a chain. The relevance of this case is determined by the large number of applications, of which, first of all, it is necessary to allocate such areas as machine learning and search for associative rules in databases. If the number of elements in each chain is two, then the dualization over the product of chains reduces in a well-known way to the construction of irreducible coverings of the Boolean matrix (the dualization of the Boolean matrix). In this paper, it is shown that in general case, when the number of elements in each chain is greater than two, the posed problem reduces to finding a subset of the set of irreducible coverings of a special Boolean matrix the size of which increases with increasing number of elements in the chain (the number of columns of the matrix grows linearly). The results of numerical experiments based on the effective "in the typical case" (for almost all variants of the problem) asymptotically optimal search for irreducible coverings of the Boolean matrix are presented. The algorithm of dualization of boolean matrix Runc-M, constructed by E. V. Djukova and P. A. Prokofjev earlier, is modified for the experiments. Runc-M is currently the world leader in counting speed. Previously, to solve the problem of dualization over the product of chains, an approach was proposed of interest mainly for the theory and is aimed at constructing incremental algorithms with quasi-polynomial time estimates "in the worst case" (E. Boros, K. Elbassioni, V. Gurvich, L. Khachiyan, and K. Makino, 2002).

Keywords: *dualization; product of partially ordered sets; chain; covering of Boolean matrix; asymptotically optimal algorithm*

DOI: 10.21469/22233792.3.4.02

References

[1] Johnson, D. S., M. Yannakakis, and C. H. Papadimitriou. 1988. On general all maximal independent sets. *Inform. Process. Lett.* 27:119–123.

*The research was supported by the Russian Foundation for Basic Research (grant 16-01-00445).

[2] Chegis, I. A., and S. V. Yablonskiy. 1958. Logicheskie sposoby kontrolya ehlektricheskih skhem [Logical ways to control electrical circuits]. *Sb. statey po matematicheskoyMIANlogike i ee primeneniyam k nekotorym voprosam kibernetiki: Tr. MIAN SSSR* [Collection of papers onmathematical logic and its applications to someMIASquestions of cybernetics: MIAS USSR Proceedings]. 51:270–360.

[3] Fredman, L., and L. Khachiyan. 1996. On the complexity of dualization of monotone disjunctive normal forms. *J. Algorithm.* 21:618–628. doi: 10.1006/jagm.1996.0062.

[4] Djukova, E. 1977. On an asymptotically optimal algorithm for constructing irredundant tests. *Sov. Math. Dokl.* 18(2):423–426.

[5] Djukova, E. 1987. The complexity of the realization of certain recognition procedures. *Comp. Math. Math. Phys.* 27(1):74–83. doi: 10.1016/0041-5553(87)90121-2.

[6] Djukova, E., and Y. Zhuravlev. 2000. Discrete analysis of feature descriptions in recognition problems of high dimensionality. *Comp. Math. Math. Phys.* 40(8):1214–1227.

[7] Djukova, E. 2004. On the implementation complexity of the realization of discrete (logical) recognition procedures. *Comp. Math. Math. Phys.* 44(3):532–541.

[8] Djukova, E., and A. Inyakin 2008. Asimptoticheski optimal'noe postroenie tupikovykh pokrytiy tselochislennoy matritsy [Asymptotically optimal irredundent tests enumeration for integer matrix]. *Matematicheskie voprosy kibernetiki* [Mathematical problems of cybernetics]. Moscow: Nauka. 17:235–246.

[9] Djukova, E. V., and P. A. Prokofjev. 2015. Asymptotically optimal dualization algorithms. *Comp. Math. Math. Phys.* 55(5):891–905. doi: 10.1134/S0965542515050103.

[10] Murakami, K., and T. Uno. 2014. Efficient algorithms for dualizing large-scale hypergraphs. *Discrete Appl. Math.* 170:83–94. doi: 10.1016/j.dam.2014.01.012.

[11] Boros, E., K. Elbassioni, V. Gurvich, L. Khachiyan, and K. Makino. 2002. Dual-bounded generating problems: All minimal integer solutions for a monotone system of linear inequalities. *SIAM J. Comput.* 31(5):1624–1643. doi: 10.1137/S0097539701388768.

[12] Andreev, A. E. 2010. Ob asimptoticheskom povedenii chisla tupikovykh testov i minimal'noy dliny testa dlya pochti vsekh tablits [On the asymptotic behavior of the number of irredundant tests and the minimal test length for almost all tables]. *Problemy kibernetiki* [Problems of cybernatics] 41:117–141.

Received August 30, 2017