О построении тупиковых покрытий булевых и целочисленных матриц

Дюкова Е. В., Инякин А. С.
djukova@ccas.ru, inyakin@ccas.ru
Москва, Вычислительный центр РАН

Задача построения тупиковых покрытий целочисленной, в частности, булевой, матрицы возникает при конструировании логических процедур распознавания и классификации и относится к числу трудных в вычислительном плане задач [1, 2, 3, 4]. Данная задача может быть также сформулирована как задача построения максимальных конъюнкций двузначной логической функции, заданной конъюнктивной нормальной формой (КНФ). Поиски эффективных алгоритмов ее решения ведутся с середины 1950-х годов.

Первостепенное значение имеет задача построения неприводимых покрытий (или тупиковых $(0, \dots, 0)$-покрытий) булевой матрицы (или задача построения максимальных конъюнкций монотонной булевой функции, заданной КНФ).

Пусть $L = (a_{ij})$, $i = 1, \dots, m$, $j = 1, \dots, n$, — булева матрица. Будем предполагать, что L не содержит строк вида $(0, \dots, 0)$. *Неприводимым* покрытием матрицы L называется набор H из r столбцов этой матрицы такой, что подматрица матрицы L, образованная столбцами набора H, не содержит строку $(0, \dots, 0)$ и содержит каждую из строк $Q_1 = (1, 0, 0, \dots, 0, 0)$, $Q_2 = (0, 1, 0, \dots, 0, 0)$, $\dots$, $Q_r = (0, 0, 0, \dots, 0, 1)$. Подматрица матрицы L, составленная из строк $Q_1, \dots, Q_r$, называется единичной подматрицей.

Пусть $P(L)$ — множество всех неприводимых покрытий матрицы L.

Как правило, число неприводимых покрытий растет экспоненциально с ростом размера матрицы, поэтому эффективность алгоритмов построения неприводимых покрытий имеет смысл оценивать в терминах полиномиальной задержки.

Будем говорить, что алгоритм, порождающий множество $P(L)$, имеет задержку (delay) d, если выполнены следующие условия: 1) алгоритм выдаст первое неприводимое покрытие, выполнив не более d элементарных операций; 2) после выдачи очередного неприводимого покрытия он либо выполнит не более d элементарных операций прежде, чем выдаст следующее неприводимое покрытие, либо остановится. Под элементарной операцией понимается просмотр одного элемента матрицы L.

Алгоритм с задержкой, ограниченной сверху полиномом от размера матрицы, назовем алгоритмом с полиномиальной задержкой.

В [1, 2, 3] рассмотрен случай, когда число строк m матрицы имеет более низкий порядок роста, чем число столбцов n, при условии, что

$n \to \infty$. Для этого случая построен асимптотически оптимальный алгоритм поиска неприводимых покрытий (алгоритм АО1). Данный алгоритм строит с задержкой, не превосходящей $O(mn)$, приближенное решение, в качестве которого рассматривается совокупность наборов столбцов, содержащих единичные подматрицы. Каждый такой набор алгоритм АО1 строит столько раз, сколько единичных подматриц он содержит. Показано, что если $m^\alpha \leqslant n \leqslant 2^{m^\beta}$, $\alpha > 1$, $\beta < 1$, то при $n \to \infty$ число шагов алгоритма, равное числу единичных подматриц, почти всегда (для почти всех матриц размера $m \times n$) асимптотически равно мощности $P(L)$. При конструировании алгоритма АО1 в качестве приближенного решения может быть рассмотрена совокупность наборов столбцов, содержащих «максимальные» единичные подматрицы. Каждый такой набор столбцов строится столько раз, сколько максимальных единичных подматриц он содержит. Тогда алгоритм будет делать меньшее число шагов и работать с задержкой, не превосходящей $O(qmn)$, здесь и далее $q = \min(m, n)$. Данная модификация алгоритма АО1 обычно используется при его реализации на ЭВМ.

В [4] построен алгоритм, основанный на переборе с задержкой, не превосходящей $O(qm^2n)$ неприводимых покрытий матрицы L (алгоритм АО2). Однако данный алгоритм строит каждый набор из $P(L)$ столько раз, сколько единичных подматриц он содержит. Из сказанного выше следует, что указанный недостаток не существенен при $m^\alpha \leqslant n \leqslant 2^{m^\beta}$, $\alpha > 1$, $\beta < 1$ (в этом случае число шагов алгоритма, равное числу единичных подматриц, порождающих неприводимые покрытия, почти всегда при $n \to \infty$ асимптотически равно мощности $P(L)$).

Отметим, что проверка на повторяемость построенного на очередном шаге набора столбцов в алгоритмах АО1 и АО2 требует просмотра не более qm элементов матрицы L.

В докладе предложен «точный» алгоритм поиска неприводимых покрытий булевой матрицы с задержкой, не превосходящей $O\left(qmn^2(m+q)\right)$. Проведено экспериментальное обоснование алгоритма на случайных матрицах. Аналогичный результат получен для задачи поиска тупиковых покрытий целочисленной матрицы. При этом использовалось сведение задачи поиска тупиковых покрытий целочисленной матрицы к задаче поиска неприводимых покрытий булевой матрицы, приведенное в [5].

Работа выполнена при поддержке проектов РФФИ № 07-01-00516, № 05-01-00495 и гранта Президента РФ по поддержке ведущих научных школ НШ № 5833.2006.1 «Алгебраические и логические методы в задачах распознавания и прогнозирования».

Литература

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

[2] *Дюкова Е. В.* Алгоритмы распознавания типа Кора: сложность реализации и метрические свойства // Распознавание, классификация, прогноз (математические методы и их применение). — М.: Наука, 1989. Вып. 2. С. 99–125.

[3] *Дюкова Е. В., Журавлёв Ю. И.* Дискретный анализ признаковых описаний в задачах распознавания большой размерности // ЖВМиМФ. — 2000. — Т. 40, №8. — С. 1264–1278.

[4] *Дюкова Е. В.* О сложности реализации дискретных (логических) процедур распознавания // ЖВМиМФ. — 2004. — Т. 44, №3. — С. 550–572.

[5] *Дюкова Е. В., Инякин А. С.* О процедурах классификации, основанных на построении покрытий классов // ЖВМиМФ. — 2003. — Т. 43, №12. — С. 1910–1921.