

DOI: 10.12737/conferencearticle_5fe01d9b660121.60953168

© 2020 Ю.В. Нагребецкая, канд. физ.-мат. наук; В.Г. Панов, канд. физ.-мат. наук

Уральский федеральный университет, Екатеринбург,

Институт промышленной экологии УрО РАН, Екатеринбург

ОЦЕНКА СЛОЖНОСТИ ЭФФЕКТИВНЫХ АЛГОРИТМОВ ПРОВЕРКИ НАЛИЧИЯ СОВМЕСТНОГО ДЕЙСТВИЯ БИНАРНЫХ ФАКТОРОВ

Приведены эффективные алгоритмы проверки наличия совместного действия k факторов в данном отклике, зависящем от n факторов, а также вычисления степеней этого совместного действия для любого k . Показано, что асимптотическая времененная сложность предлагаемых алгоритмов не превосходит квадрата от размера входных данных, представляющих отклик f .

Ключевые слова: булева функция, булев куб, эффективный алгоритм, временная сложность алгоритма, бинарная теория достаточных причин, отклик.

© 2020 J.V. Nagrebetskaya, PhD, V.G. Panov, PhD

Ural Federal University, Ekaterinburg; Institute of Industrial Ecology, UrB of RAS, Ekaterinburg

ESTIMATION OF COMPLEXITY OF EFFECTIVE ALGORITHMS FOR CHECKING A PRESENCE OF JOINT ACTION OF BINARY FACTORS

Effective algorithms are provided for checking presence of joint action of k factors in a given outcome which depends on n factors ($k < n$) and for calculation of degrees of that joint action for any k . It is demonstrated that asymptotic time complexity of the proposed algorithms does not exceed square of the input data size representing the given outcome.

Key words: Boolean functions, Boolean cube, effective algorithm, computational complexity, binary sufficient cause theory, outcome.

Введение. Булева модель теории достаточных причин, предложенная в работах [1–8], позволяет рассматривать эту эпидемиологическую концепцию как приложения общей теории конечных булевых алгебр и булевых функций [9, 10]. При этом помимо прояснения и формализации основных понятий этой концепции возникает необходимость разработки эффективных алгоритмов [11, 12], вычисляющих различные величины обсуждаемой теории. При этом важно не только предложить эти алгоритмы, но и оценить их вычислительную сложность [11, 12], поскольку размер входных данных для этих алгоритмов в булевом формализме теории достаточных причин достаточно высок.

Ниже мы рассмотрим оценки асимптотической временной вычислительной сложности проверки наличия совместного действия факторов в отклике и нахождения степени этого совместного действия [5, 6, 8], применяемых в задаче классификации совместного действия факторов в бинарной теории достаточных причин.

Предварительно заметим, что речь будет идти только об *асимптотической сложности*, которая обычно используется в теории сложности алгоритмов [11, 12], что связано с размером входных данных обсуждаемой задачи, зависящим от числа факторов экспоненциально. Действительно, булев куб B^n имеет 2^n вершин, а любой двухуровневый отклик f от n (булевых) переменных полностью определяется своим носителем C_f , причем $|C_f| \leq 2^n$ [7]. Так как вершина булева куба определяется бинарным вектором длины n , то размер входных данных, задающих отклик f , в худшем случае равен $N = n \cdot 2^n$. Следовательно, уже при сравнительно небольшом числе факторов объем исходных данных становится слишком большим, что делает необходимым построение эффективных алгоритмов вычисления различных конструкций этой теории, причем один из параметров оценки алгоритма как эффективного является минимизация его вычислительной сложности. Кроме

того, следует отметить, что имеющиеся модели теории достаточных причин как в формальном [1-8], так и в полуформальном изложении [14-16] допускают *произвольное* число факторов. Поэтому уже с точки зрения полноты этой теории рассмотрение всех вопросов, возникающих в контексте того или иного формализма, совершенно естественно и необходимо.

Одно из важных общих понятий булевой модели теории достаточных причин – понятие совместного действия n факторов в отклике, зависящем от всех этих n факторов, введено в работах [5,7]. В работе [7] также был приведён эффективный алгоритм проверки по данному отклику наличия в нём совместного действия всех его факторов, а также приведена оценка сверху временной сложности этого алгоритма величиной $O(4^n \cdot n)$.

Основные результаты

В работах [6, 8] было предложено понятие взаимодействия k ($2 \leq k \leq n$) факторов в отклике, зависящем от n факторов, которое обобщает понятия взаимодействия n факторов в отклике, зависящем от n факторов. С помощью критерия совместного действия k факторов, предложенного в [6], можно доказать аналогичную оценку временной сложности эффективного алгоритма проверки наличия этого действия в данном отклике. Формулировка этого критерия (см. Теорему 1 ниже) использует представления отклика f в виде неориентированного графа Γ_f , который является подграфом булева куба B^n с вершинами из носителя C_f .

Теорема 1 [6]. В отклике f , зависящем от n факторов, имеется взаимодействие k факторов тогда и только тогда, когда минимальная степень вершин в графе Γ_f не превосходит $n - k$.

Теорема 2. Для любого k ($2 \leq k \leq n$) существует эффективный алгоритм, проверяющий наличие совместного действия k факторов в отклике, зависящем от n факторов, причем временная сложность этого алгоритма имеет порядок не более чем $O(4^n \cdot n)$.

Доказательство. Каждый автоморфизм t булева куба B^n может быть продолжен на свободную булеву алгебру всех булевых функций от переменных x . При этом булевой функции f соответствует булева функция $t(f)$, чей носитель $C_{t(f)}$ равен образу $t(C_f)$ носителя C_f функции f . Пусть $\gamma \in \Gamma_f$ – произвольная вершина графа Γ_f . Рассмотрим автоморфизм $t_\gamma(x)$ булева куба B^n , определяемый равенством $t_\gamma(x) = x \oplus \gamma$, где \oplus – сложение по модулю 2. Любой автоморфизм булева куба является изометрией этого куба как метрического пространства с метрикой-расстоянием Хэмминга. Поэтому степень вершины γ в графе Γ_f равна степени вершины $\mathbf{0}$ в графе $\Gamma_{t_\gamma(f)}$, вершинно-порожденном множеством $C_{t_\gamma(f)}$. Степень вершины $\mathbf{0}$ в графе $\Gamma_{t_\gamma(f)}$ равна числу вершин единичного веса Хэмминга в множестве $C_{t_\gamma(f)}$. Получение списка таких вершин и подсчёт их числа имеют сложность не более чем $O(2^n \cdot n)$. Следовательно, вычисление степени любой вершины в графе Γ_f имеет временнюю сложность не более чем $O(2^n \cdot n)$, а вычисление минимальной степени вершин в этом графе – не более чем $O(4^n \cdot n)$.

Таким образом, оценка $O(4^n \cdot n)$ экспоненциальна по числу факторов n , но менее чем квадратична по наибольшему размеру входных данных $N = 2^n \cdot n$.

Для исследования совместного действия факторов в данном отклике в работах [5,7] было введено понятие степени μ_f совместного действия n факторов в отклике f . В работе [7] также предложен эффективный алгоритм нахождения μ_f и показано, что верхняя оценка сложности этого алгоритма равна $O(4^n \cdot n)$.

В работе [6] было введено обобщение этого понятия – степень $\mu_{f,k}$ совместного действия k факторов в отклике f , зависящем от n факторов, а в работе [8] рассмотрена зависимость степени $\mu_{f,k}$ от k . Следующее утверждение даёт оценку сложности вычисления набора $M_f = (\mu_{f,1}, \dots, \mu_{f,n})$ степеней совместного действия для данного отклика f .

Теорема 3. Существует эффективный алгоритм вычисления набора M_f степеней совместного действия данного отклика f , причем сложность этого алгоритма оценивается сверху величиной порядка $O(n^2 4^n) = O(N^2)$, где N , как и раньше, максимальный размер входных данных, $N = 2^n \cdot n$.

Доказательство. Рассмотрим булев куб B^n как решёточно-упорядоченное множество $\langle B^n, \leq \rangle$ относительно покомпонентного сравнения ‘ \leq ’ бинарных векторов [10], которое можно представить ориентированным графом R . В графе R для каждого $k \in \{1, \dots, n\}$ выделены списки U_k лексикографически упорядоченных вершин веса k . На формирование графа R потребуется времени (элементарных шагов вычислительных операций) не более чем $O(4^n \cdot n)$. Далее для каждого набора $\alpha \in B^n$ производим следующие шаги:

- (1) формируем список лексикографически упорядоченных вершин множества C_g , где $g = t_\alpha(f), t_\alpha(x) = x \oplus \alpha$ из списка вершин множества C_f .
- (2) полученный список распределяем по спискам V_k , $k \in \{1, \dots, n\}$, вершин веса k , для которых сохраняется лексикографическое упорядочение;
- (3) при помощи списков V_k из графа R получаем помеченный ориентированный граф R_g , в котором каждая вершина γ из списка U_k помечается числом $l(\gamma) = 1$, если $\gamma \in C_g$ и $l(\gamma) = 0$ в противном случае. Таким образом, граф R_g представляет частично упорядоченное множество $\langle C_g, \leq \rangle$.

Шаги (1)-(3) в целом займут времени не более чем $O(2^n \cdot n)$.

(4) для каждой вершины γ графа R_g из списка U_k вычисляем метку $m(\gamma)$ с помощью метки $l(\gamma)$ и меток $m(\gamma'), l(\gamma')$ для k вершин γ' из списка U_{k-1} , покрываемых вершиной γ в $\langle B^n, \leq \rangle$ (считая, что $m(\emptyset) = 0$);

(5) для каждого $k \in \{1, \dots, n\}$ находим m_k – максимальное значение $m(\gamma)$ среди всех вершин γ графа R_g из списка U_k .

На этапы (4)-(5) потребуется время не более чем $O(2^n \cdot n \cdot \log n)$. Шаги (1)-(5) повторяются для каждого набора $\alpha \in B^n$ и поэтому в целом выполняемые действия потребуют времени не более чем $O(4^n \cdot n \cdot \log n)$, что не превосходит $O(n^2 4^n)$.

На последнем шаге для каждого k находим максимальное значение среди всех m_k , которые вычисляются на этапе (5) для всех вершин $\alpha \in B^k$. Можно доказать, что полученное значение равно степени $\mu_{f,k}$.

Таким образом, в целом, описанный динамический алгоритм имеет временную сложность не более чем $O(n^2 4^n) = O(N^2)$.

ЛИТЕРАТУРА

1. Панов В. Г., Нагребецкая Ю. В. Алгебраическая трактовка двухфакторной теории достаточных причин // Труды СПИИРАН. 2013. Т. 3(26). С. 277–296.
2. Панов В. Г., Нагребецкая Ю. В. Алгебраическая классификация совместного действия n бинарных факторов // Материалы IX междунар. конф. «Системный анализ в медицине». Благовещенск. 2015. С. 31–34.
3. Panov V.G. and Nagrebetskaya J.V. Boolean algebras and classification of interactions in sufficient-component cause model // Int. J. Pure Appl. Math. 2015. V. 98(2). P. 239–259.
4. VanderWeele T.J., Robins J.M. A theory of sufficient cause interactions // COBRA Preprint Series. 2006. Paper 13.
5. Нагребецкая Ю.В., Панов В.Г. Степень взаимодействия бинарных факторов в теории достаточных причин // Материалы XIII междунар. конф. "Системный анализ в медицине". Благовещенск, 2019. С. 31–34.
6. Нагребецкая Ю. В., Панов В. Г. Обобщение понятия взаимодействия n факторов в теории достаточных причин и его свойства // Материалы XIII междунар. конф. "Системный анализ в медицине". Благовещенск, 2019. С. 35–38.
7. Nagrebetskaya J.V., Panov V.G. Joint action of binary factors in the sufficient causes theory and its classification // Int. J Innov. Tech.&Exploring Eng. V.9(1), 2019. P. 2146–2153.
8. Яблонский С. В. Введение в дискретную математику. М.: Наука, 1986.
9. Лидл Р., Пильц Г. Прикладная абстрактная алгебра. Екатеринбург: Изд-во Уральского ун-та, 1996.
10. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982.
11. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. М.: Мир, 1979.
12. Rothman K. Causes // Am. J. Epidemiology. 1976. V. 104(6). P. 587–592.
13. Miettinen O. S. Causal and preventive interdependence: Elementary principles // Scand. J. Work. Environ. Health. 1982. V. 8. P. 159–168.
14. VanderWeele T. J., Robins J. M. The identification of synergism in the sufficient-component-cause framework // Epidemiology. 2007. V. 18(3). P. 329–339.
15. VanderWeele T. J., Richardson T. S. General theory for interactions in sufficient cause models with dichotomous exposures // Ann. Statistics. 2012. V. 40. P. 2128–2161.

E-mail: I.V.Nagrebetskaya@urfu.ru; vpanov@ecko.uran.ru