ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 001.5:519.17

С. О. Кузнецов

ИНТЕРПРЕТАЦИЯ НА ГРАФАХ И СЛОЖНОСТНЫЕ ХАРАКТЕРИСТИКИ ЗАДАЧ ПОИСКА ЗАКОНОМЕРНОСТЕЙ ОПРЕДЕЛЕННОГО ВИДА

Предложена интерпретация на графах задачи поиска закономерностей с помощью ДСМ-метода автоматического порождения гипотез. Эта интерпретация позволяет устанавливать связь между поиском ДСМ-гипотез и изучением некоторых классических комбинаторных объектов (двудольных графов и полных двудольных подграфов в них). Тем самым устанавливается связь и с некоторыми другими системами поиска закономерностей. Устанавливается ряд фактов относительно алгоритмической сложности (полиномиальная вычислимость, NP-полнота, #P-полнота) некоторых задач поиска закономерностей указанного вида (и, следовательно, поиска подграфов некоторого вида в двудольных графах).

Одним из способов задания отношения сходства объектов как основы поиска закономерностей в данных может служить введение некоторой операции «выделения общей части» объектов. Такое понимание сходства лежит в основе целого ряда интеллектуальных систем и систем распознавания, в том числе в основе ДСМ-метода автоматического порождения гипотез [1—4].

В частности, когда объекты представлены множествами, то такой операцией может служить операция пересечения, а поиск сходства сводится к поиску всевозможных пересечений исходных множеств, удовлетворяющих некоторым дополнительным ограничениям, обусловленным особенностью решающих средств конкретной интеллектуальной системы (или системы распознавания). В ДСМ-методе, например, решающие средства основаны на особых признаках и правилах правдоподобного вывода, сформулированных на специальном языке, являющемся расширением языка логики предикатов первого порядка.

Первый вопрос, который будет рассмотрен в работе, касается интерпретации на графах задачи поиска закономерностей некоторого рода. Подобно тому, как геометрическая интерпретация (на булевых кубах) задачи минимизации булевых функций [5] помогает в ряде случаев быстро находить ответы на некоторые вопросы в указанной области, интерпретация на графах поиска закономерностей, о которых будет идти речь ниже, позволяет легко отвечать на некоторые вопросы, связанные с построением эффективных алгоритмов, находящих эти закономерности.

В работе используется упрощенное понимание ДСМ-гипотезы (закономерности), связанное с использованием лишь примеров объектов, обладающих определенным свойством (в отсутствии контрпримеров) одноэлементного множества свойств, и самого простого решающего предиката — предиката простого сходства Ma^+. Это упрощение имеет двоякое обоснование.

Во-первых, при таком упрощении видна связь задачи поиска ДСМ-гипотез с задачами поиска гипотез в других системах (см. [6], а также список литературы, помещенный в этой книге). Упрощение выявляет комбинаторное ядро интеллектуальных систем и систем распознавания некоторого рода, позволяя интерпретировать результаты относительно алгоритмической сложности решения некоторых задач ДСМ-метода как результаты относительно алгоритмических характеристик классических комбинаторных объектов — двудольных графов и бинарных матриц.

Во-вторых, негативные результаты (об NP- и $\#P$-полноте некоторых задач) в упрощенной постановке влекут аналогичные результаты в более общих постановках, использующих контрпримеры, более сильные решающие предикаты и более сложные структуры данных.

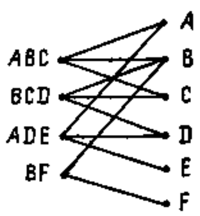
Итак, пусть заданы некоторые конечные множества U^1 и $\Omega\subseteq 2^{U^1}$. Пару $\langle U^1, \Omega\rangle$ будем называть входными данными ДСМ-задачи. (Множество U^1 соответствует структурным элементам, а множество Ω — объектам, обладающим некоторым интересующим нас свойством A.) ДСМ-гипотезой (о причине свойства A) назовем пару $\langle H\{X_1, \dots, X_l\}\rangle$, где $H=X_1\cap\dots\cap X_l$, $X_i\in\Omega$ для $i=\overline{1, l}$ и для любого X: $X\in\Omega\setminus\{X_1, \dots, X_l\}$ выполняется $H\cap X\neq H$.

Обозначим также $h=|H|$.

Итак, сопоставим множествам $U^1\Omega$ двудольный граф

$G(\mathrm{U}^1\Omega)$ следующим образом: в правой доле каждая вершина взаимно-однозначно соответствует некоторому элементу из U^1, а в левой доле — некоторому множеству из Ω. Вершины i и j связаны в G ребром тогда и только тогда, когда множество $X_i \in \Omega$ содержит элемент $s_j \in \mathrm{U}^1$.

Пример:



$\mathrm{U}^1 = \{A, B, C, D, E, F\}$,

$\Omega = \{ABC, BCD, ADE, BF\}$.

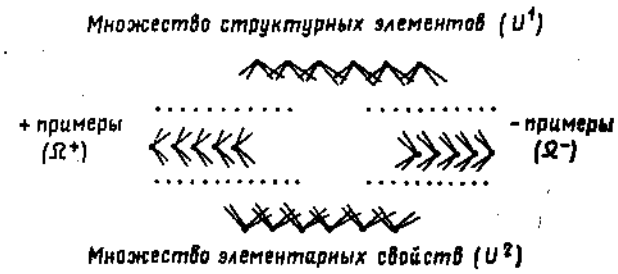
$\mathscr{H}$ — множество всех гипотез: $\mathscr{H} = \{\langle BC, \{ABC, BCD\}\rangle, \langle A, \{ABC, ADE\}\rangle, \langle B, \{ABC, BCD, BF\}\rangle$.

Нетрудно видеть, что каждая гипотеза соответствует максимальному по вложению полному двудольному подграфу R графа G, у которого в левой доле не меньше двух вершин, т. е. такому подграфу вида $(V^1 \cup V^2, E)$, $E = V^1 \times V^2$, $|V^1| \geqslant 2$, что при добавлении к любому V^l, $l \in \{1, 2\}$ какой-либо вершины v из той же доли G индуцированный подграф на вершинах $V^1 \bigcup \{G\} \cup V^2$ не будет полным, т. е. для соответствующего множества $E' : E' \subset (V^1 \cup \{v\}) \times V^2$ (если в качестве V^l взято V^1).

Так, подграф, индуцированный вершинами ABC, BCD, B, C, является максимальным по вложению полным двудольным подграфом графа G из примера: добавление какой-либо другой вершины нарушает его полноту (например, при добавлении вершины BF, вершина C оказывается не связанной с BF). И обратно, пусть имеется произвольный двудольный граф $B = (V^1 \cup V^2, E)$, $V^1 = \{1, \ldots, r\}$, $V^2 = \{b_1, \ldots, b_t\}$. Сопоставим правую долю B (вершины из V^2) множеству U^1, а левую (вершины V^2) — множеству Ω некоторой задачи о ДСМ-гипотезах. Для каждой вершины i из V^1 образуем новую вершину a_i и добавим ее к множеству V^2, и соединим ребром только с a_i (это нужно для того, чтобы в Ω не было повторяющихся множеств). Теперь каждой вершине i из V^1 сопоставим множество $w_i = \{a_i, b_{j_1}, \ldots, b_{j_k}\}$, где $\{b_{j_1}, \ldots, b_{j_k}\}$ в точности множество тех вершин из V^2, которые связаны с вершиной i. Теперь двудольному графу $(V^1 \cup V^2 \cup \{a_1, \ldots, a_r\}, E \cup \{(1, a_1), \ldots, (r, a_2)\})$ соответствуют входные условия задачи о ДСМ-гипотезах с $\mathrm{U}^1 = \{a_1, \ldots, a_r, b_1, \ldots, b_t\}$ и $\Omega = \{w_1, \ldots, w_r\}$, а каждому максимальному по вложению полному двудольному подграфу R графа B, у которого в левой доле не меньше двух вершин, соответствует ДСМ-гипотеза для входных данных U^1, Ω. Сведе́ние в обе стороны осуществляется за линейное время. Поэтому верна следующая

Лемма 1. По входным данным $\langle \mathrm{U}^1, \Omega \rangle$ ДСМ-задачи за линейное от $|\mathrm{U}^1| \cdot |\Omega|$ время можно построить двудольный граф, максимальные полные по вложению двудольные подграфы (с числом вершин в одной из долей не меньшим двум), которые взаимно-однозначно соответствуют гипотезам ДСМ-задачи. И наоборот, по любому двудольному графу $B = (V^1 \cup V^2, E)$ за линейное время можно построить двудольный граф $B' = ((V^1 \cup V) \cup V^2, E)$, по которому, в свою очередь, можно построить входные данные $(\mathrm{U}^1(B'), \Omega(B'))$ ДСМ-задачи, причем гипотезы взаимно-однозначно соответствуют полным максимальным по вложению двудольным подграфам, содержащим не менее двух вершин из V^2.

Более полная постановка задачи нахождения ДСМ-гипотез, подразумевающая наличие в базе данных объектов, обладающих данным свойством (+-примеров), и объектов, не обладающих данным свойством (—-примеров), также позволяющая рассматривать не только элементарные свойства, но и множества различных элементарных свойств, может интерпретироваться как поиск некоторых подграфов в четырехдольном графе Q следующего вида:



с множеством ребер $E \subseteq \mathrm{U}^1 \times \Omega^+ \cup \mathrm{U}^1 \times \Omega^- \cup \mathrm{U}^2 \times \Omega^+ \cup \mathrm{U}^2 \times \Omega^-$. Более точно, существованию гипотезы относительно наличия свойства $\mathbb{W} \subseteq \mathrm{U}^2$ в этом случае (например, для ситуации, когда используется так называемый простой предикат сходства Ma^+ [1]) будет соответствовать существование в графе Q максимального по вложению полного двудольного подграфа, индуцированного вершинами из U^1, Ω^+ (пусть это будет подграф на множествах вершин V^3 и V^4, $V^3 \subset \mathrm{U}^1$, $V^4 \subset \Omega^+$), такого, что все вершины из V^4 соединены ребрами со всеми вершинами из $\mathbb{W}$. При этом в Q не должно существовать максимального по вложению полного двудольного подграфа, индуцированного вершинами из U^1, Ω^- (пусть это будет подграф на множествах вершин V^5 и V^6, $V^5 \subset \mathrm{U}^1$, $V^6 \subset \Omega^-$), такого, что все вершины из V^6 соединены ребрами со всеми вершинами из $\mathbb{W}$.

Вернемся, однако, к упрощенной постановке ДСМ-задачи.

Другой возможной интерпретацией задачи нахождения ДСМ-гипотез может быть задача нахождения экстремальных единичных подматриц бинарных матриц. Об эквивалентности этой задачи задаче поиска полных двудольных подграфов двудольного графа упоминается, например, в монографии [7] в связи с проблемой Царанкевича* (см. также [6]).

Рассмотрим теперь некоторые вопросы, связанные со сложностью порождения гипотез, и покажем, как интерпретация на графах помогает решать некоторые из них. Даже при указанных выше ограничениях задача порождения всех гипотез может иметь экспоненциальную сложность. В самом деле, при $\overline{\mathrm{U}}^1 = \{a_1, \ldots, a_n\}$ и $\overline{\Omega} = \{\mathbb{W}_1, \ldots, \mathbb{W}_n\}$, где $\mathbb{W}_i = \mathrm{U}^1 \setminus \{a_i\}$, число всех возможных ненулевых пересечений множеств из Ω составляет $2^n - n - 2$ [8]. Эта величина является точной верхней оценкой числа пересечений. В дальнейшем условия задачи нахождения ДСМ-гипотез, представленные $\overline{\mathrm{U}}^1$ и $\overline{\Omega}$, будем называть n-базисом и обозначать $\mathscr{B}(\overline{\mathrm{U}}^1)$. Экспоненциальная верхняя оценка числа гипотез говорит о том, что задача нахождения всех гипотез может быть в ряде случаев трудновыполнимой, поэтому целесообразна постановка вопроса о «быстром» нахождении

---

* Проблема Царанкевича заключается в оценке числа κ, для которого любая $n \times n$ — бинарная матрица, содержащая κ единичных элементов, содержит единичную подматрицу размера $\alpha \times \alpha$.

«хороших» гипотез. Показателем качества гипотезы $\langle H, \{X_1, \ldots, X_l\}\rangle$ могут служить, например, следующие функционалы:

1: h — мощность H (как некоторая мера информативности гипотезы);

2: l — число множеств, образующих гипотезу в качестве пересечения (как некоторая мера надежности гипотезы);

3: $h+l$ — интегральная мера информативности и надежности.

В связи с выбором функционала качества f гипотезы встают следующие проблемы (для функционалов 1 и 2 они были поставлены в работе [8]):

Задача 1. Существует ли гипотеза, для которой значение функционала f не меньше k $(f \geqslant k)$?

Задача 2. Существует ли гипотеза, для которой значение функционала f не больше k $(f \leqslant k)$?

Задача 3. Существует ли гипотеза со значением функционала в точности k $(f=k)$?

На двудольных графах задачи 1—3 интерпретируются как задачи поиска максимальных по вложению полных подграфов с ограничениями на размер левой доли (функционал l), правой доли (функционал h) и всего подграфа (функционал $h+l$).

В работе [8] было показано, что задачи 1 и 2 для функционалов h и l решаются за время $O(n^3)$.

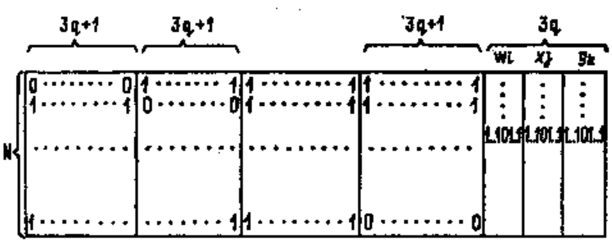
Интерпретация задачи поиска ДСМ-гипотез на графах позволяет моментально установить алгоритмическую эквивалентность задач 1—3 с функционалом h, а задач 1—3 — с функционалом l. (Этот факт двойственности был установлен в работе [8].) В самом деле, алгоритму нахождения экстремальных полных двудольных подграфов безразлично, накладывается ли ограничение на размер левой доли или на размер правой (с точностью до поправки в $O(n^2)$ во времени, из-за проверки на экстремальные полные подграфы с единственной вершиной в одной из долей — в пересечении не может участвовать менее двух множеств).

Сложность решения задачи 3 для функционалов h, l не была известна, сложность решения задач 1—3 для функционала $h+l$ ранее не исследовалась.

Сначала покажем, что задача 3 с функционалом h (а, следовательно, и с функционалом l) NP-полна. С этой целью докажем полиномиальную сводимость к ней задачи «трехмерное сочетание» (3-C) [9]. Напомним формулировку (3-C): дано множество $M \subseteq W \times X \times Y$, где W, X и Y — непересекающиеся множества, содержащие одинаковое число элементов q. Верно ли, что M содержит трехмерное сочетание, т. е. подмножество $M' \subseteq M$, такое, что $|M'|=q$, и никакие два разных элемента M' не имеют ни одной равной координаты?

Теорема 2*. Задача 3 с функционалом h является NP-полной.

Доказательство. Пусть нам дана индивидуальная задача из (3-C) с параметрами M $(|M|=N)$, W, X, Y $(|W|=|X|=|Y|=q)$. Построим следующую бинарную матрицу $\mathscr{D}$



* Доказательство теоремы 2 одновременно и независимо было получено А. А. Разборовым. Впоследствии еще одно доказательство было предложено М. И. Забежайло.

В правой части матрицы — $3q$ правых позиций — каждая строка соответствует элементу $m \in M$ так, что нуль в столбцах w_i, x_j, y_k означает наличие соответствующей компоненты, а единица — отсутствие. В левой части матрицы — $N(3q+1)$-левых позиций в i-й строке нули стоят в позициях с $(i-1)(3q+1)+1$ по $i \cdot (3q+1)$, а остальные позиции заполнены единицами. Если задача (3-C) имеет решение, то в матрице $\mathscr{D}$ найдется q строк, таких, что в правой части матрицы в каждом столбце будет стоять по одному нулю, и произведение всех q строк будет строкой, содержащей ровно $q(3q+1)+3q$ нулей и, соответственно, $N(3q+1)+3q-(q \cdot 3q+1)+3q=(N-q) \cdot (3q+1)$ единиц ($3q$ нулей даст правая часть, а $q \cdot (3q+1)$ нулей — левая часть матрицы). Таким образом, каждому решению задачи 3-C с параметрами q, N соответствует решение задачи I с матрицей вида $\mathscr{D}$ и параметрами $N(3q+1)+3q$ (максимальная длина строки), N (число строк), $q(3q+1)+3q$ (интересующее нас число единиц булевого вектора, являющегося произведением некоторого подмножества строк).

Наоборот, пусть в построенной матрице $\mathscr{D}$ произведение некоторого подмножества строк M_2 содержит $q(3q+1)+3q$ нулей. Число строк в M_2 не может быть меньше q, так как тогда левая часть матрицы дала бы не более $(q-1) \cdot (3q+1)$, а правая не может быть более $3q$, и в сумме $(q-1) \cdot (3q+1)+3q<q(3q+1)+3q$. Число строк в M_2 не может быть больше q, так как тогда левая часть матрицы дала бы в произведении не менее $(q+1)(3q+1)$ нулей, а $(q+1)(3q+1)>q(3q+1)+3q$. Таким образом, число строк в M_2 ровно q, число нулей в левой части суммы строк из M_2 должно быть $q(3q+1)$. Значит, число нулей в правой части произведения строк есть $q(3q+1)+3q-q(3q+1)$, а это возможно, когда в каждом столбце правой части матрицы $\mathscr{D}$, соответствующей строкам из M_2, стоит ровно один нуль, т. е. W содержит трехмерное сочетание, соответствующее множеству правых частей строк из M_2.

Следствие. Задача 3 с функционалом l является NP-полной. Путем сведения задачи (3-C) к задаче 3 для функционала $h+l$ покажем, что верна следующая

Теорема 3. Задача 3 для функционала $h+l$ NP-полна.

Доказательство. По индивидуальной задаче (3-C) построим матрицу $\mathscr{D}'$, устроенную так же, как матрица $\mathscr{D}$, с тем лишь отличием, что каждый участок левой подматрицы с нулями в определенной строке имеет ширину не $3q+1$, а q^3+1.

Если для задачи (3-C), представленной матрицей $\mathscr{D}'$, есть решение, т. е. трехмерное сочетание W', то произведение соответствующих q строк в правой части ($3q$ позиций) не будет содержать ни одной единицы, а в левой — будет содержать $N(q^3+1)-q(q^3+1)$ единиц. Сумма числа строк и числа единиц в их произведении будет $q+(N-q) \cdot (q^3+1)$.

И наоборот, пусть для некоторого подмножества M_3 множества всех строк матрицы $\mathscr{D}'$ сумма числа строк и числа единиц в их произведении будет $q+(N-q) \cdot (q^3+1)$. При этом число строк $k=|M_3|$ не может быть меньше q, так как в противном случае число единиц в произведении строк было бы не меньше $(N-q+1)(q^3+1)$: так как $k \leqslant q-1$, то $k+(N-k) \cdot (q^3+1) \geqslant (N-q+1)(q^3+1)>q+(N-q)(q^3+1)$. Число строк k не может быть большим q, так как тогда бы число единиц в произведении строк было бы не больше $(N-q-1)(q^3+1)$: так как $q+1 \leqslant k \leqslant N \leqslant q^3$, то $k+(N-q-1)(q^3+1) \leqslant k+(N-q-1) \cdot (q^3+1)=(k-q^3-1)+(N-q) \cdot (q^3+1)<q+(N-q) \cdot (q^3+1)$.

В любом случае (при $k>q$ или $k<q$) сумма не могла бы совпадать с $q+(N-q) \cdot (q^3+1)$. Значит, сумма может быть такой только при числе строк q. При этом $(N-q) \cdot (q^3+1)$ единиц должно приходиться на левую часть ($N \cdot (q^3+1)$ позиций) произведения строк, а сумма числа строк k и числа единиц n_e правой части про-

изведения ($3q$ позиций) есть $k+n_e$. Так как $k=q$, то $n_e=0$, а это возможно только тогда, когда в каждом столбце правой части подматрицы $\mathscr{D}'$, соответствующей строкам из M_3, стоит ровно по одному нулю. Значит, M_3 соответствует трехмерному сочетанию.

Построение алгоритма для решения задачи 1 для функционала $h+l$ сводится к построению алгоритма, находящего размер максимального по числу вершин полного двудольного подграфа двудольного графа.

Назовем слабым дополнением двудольного графа G двудольный граф $\bar{G}$, у которого вершины совпадают с вершинами G, а пара вершин из разных долей соединена ребром тогда и только тогда, когда она не соединена ребром в графе G. Максимальному по числу вершин полному двудольному подграфу в графе G будет соответствовать максимальное независимое (несвязное) множество вершин в $\bar{G}$. Одним из следствий теоремы [10, с. 119], доказанной Кёнигом, является утверждение о том, что в произвольном двудольном графе $\bar{G}$ число ребер в максимальном паросочетании и размер максимального независимого множества в сумме дают число вершин графа $\bar{G}$.

Максимальное паросочетание в двудольном графе может быть найдено за время $O(n^{2,5})$ с помощью алгоритма Карпа и Хопкрофта (см. [13], а также, например, [12, с. 512]). Следовательно, и ответ для задачи 1 с функционалом $h+l$ находится полиномиально быстро. Указание на быстроразрешимость задачи о размере максимального полного двудольного подграфа двудольного графа сведением ее к задаче о максимальном паросочетании имеется в книге [9, с. 244]. Описание полиномиального $O(n^3)$ алгоритма нахождения максимального полного двудольного подграфа двудольного графа содержится в работе [11], что позволяет быстро находить наилучшие в смысле функционала $h+l$ гипотезы.

В традиционных постановках задачи распознавания, связанных с нахождением оптимальной в некотором смысле функции, «покрывающей» множество исходных значений, бывает достаточно найти одну такую функцию (гипотезу). В ДСМ-методе ищутся по возможности либо все гипотезы, либо те из них, которые удовлетворяют ограничениям на значения каких-либо функционалов. Следовательно, важным вопросом при исследовании алгоритмических свойств ДСМ-метода являются вопросы о числе гипотез, и о сложности алгоритма его нахождения. До последнего времени не удавалось получить нетривиальных оценок этого числа при заданных U^1, Ω общего вида. Попытаемся доказать $\#P$-полноту этой задачи и тем самым обосновать трудность получения интересных оценок такого рода.

Понятие $\#P$-полноты было введено в работах [14, 15] в связи с исследованием трудности перечислительных задач, т. е. тех, в которых требуется указать число решений задачи распознавания [9]. Будем использовать определение, использующее понятие недетерминированной машины Тьюринга, как это делается в работе [14].

Определение. Считающей машиной Тьюринга (СМТ) называется недетерминированная машина Тьюринга, дополнительно печатающая на особой ленте двоичную запись числа решений индивидуальной задачи T. Временна́я сложность СМТ есть $g(n)$, если самое долгое принимающее вычисление СМТ для задачи T на всех индивидуальных задачах размера n есть $g(n)$.

Определение. $\#P$ есть класс перечислительных задач, вычисляемых СМТ за полиномиальное время.

Определение. Перечислительная задача является $\#P$-полной, если к ней сводима, по Тьюрингу, любая задача из класса $\#P$.

В работе [16] были доказаны результаты о $\#P$-полноте большого числа перечислительных задач, соответствующих известным NP-полным задачам, в частности, задачам о клике, гамильтоновом цикле, выполнимости 3-КНФ, 3-сочетании и др. В работе [16] использовалась так называемая «консервативная» сводимость $\propto_{pa}$, при которой $A \propto_{pa} B$ подразумевает, что каждое решение задачи A взаимно-однозначно соответствует решению задачи B и, тем самым, число решений совпадает, а задачи нахождения этих чисел полиномиально эквивалентны. В работе [14] была доказана $\#P$-полнота задачи вычисления перманента матрицы, что позволило доказать $\#P$-полноту ряда перечислительных задач, для которых соответствующие задачи распознавания разрешимы за полиномиальное время.

В ДСМ-методе задачей такого рода является задача о числе всех ДСМ-гипотез для входных данных U^1 и Ω.

Полиномиальность алгоритма соответствующей задачи распознавания: «существует ли гипотеза при данных $\langle U^1, \Omega\rangle$» доказывается тривиально (см., например, [8]).

Докажем $\#P$-полноту задачи о числе всех гипотез, сведя к ней задачу о выполнимости монотонной 2-КНФ. $\#P$-полнота последней задачи была установлена в [15]. Приведем формулировку задачи о 2-КНФ.

Дано: множество переменных $Y=\{y_1, \ldots, y_n\}$ и конъюнкция $\mathscr{F}(Y)=C_1\wedge\ldots\wedge C_r$, где C_i имеет вид $C_i=(y_{i_1}\vee y_{i_2})$, $y_{i_j}\in Y$.

Найти: число булевых наборов, выполняющих $\mathscr{F}(Y)$.

Теорема 4. Задача «найти число всех гипотез» $\#P$ — полна.

Доказательство. Покажем, что, умея находить число гипотез, мы сможем найти число булевых наборов, выполняющих $\neg\mathscr{F}(Y)$: $\neg\mathscr{F}(Y)=\neg(C_1\wedge\ldots\wedge C_r)=$ $=\neg C_1\vee\ldots\vee\neg C_r=d_1\vee\ldots\vee d_r=D(Y)$, где $d_i=$ $=\bar{y}_{i_1}\cdot\bar{y}_{i_2}$. Сопоставим множеству переменных Y множество $U^1=\{a_1, \ldots, a_n\}$, причем каждой переменной y_j будет взаимнооднозначно соответствовать элемент a_j. Дизъюнктивному члену d_i дизъюнкции $D(Y)$ сопоставим множество $\mathscr{A}_{i_1i_2}=U^1\setminus\{a_{i_1}, a_{i_2}\}$. Покажем, что каждому выполняющему набору функции $D(Y)$ можно взаимно-однозначно сопоставить множество, являющееся пересечением некоторых пар индексов, соответствующих индексам переменных в конъюнкции $\mathscr{F}(Y)$. Возьмем произвольный выполняющий $D(Y)$ набор $\alpha=(\alpha_1, \ldots, \alpha_n)$. Положим $A_\alpha=\{a_i\in U^1 \mid \alpha_i=1\}$. Очевидно, что A_α принадлежит множеству пересечений, порождаемых $U^1 \bigcup\limits_{(i_1i_2)\in L} (\mathscr{B}(\mathscr{A}_{i_1i_2})\cup\{\mathscr{A}_{i_1i_2}\})=\Omega$. Обратно, пусть множество A_β порождено пересечением множеств из Ω, и для некоторых $t_1, \ldots, t_p$ $A_\beta=U^1\setminus\{a_{t_1}, \ldots, a_{t_p}\}$. Выделим множество $\mathscr{K}_\beta=\{\mathscr{A}_{i_1i_2} \mid A_\beta\subseteq\mathscr{A}_{i_1i_2}\}$. Образуем булев набор, у которого на i-м месте стоит нуль, если $i\in\{t_1, \ldots, t_p\}$. Остальные позиции сделаем единичными. Этот набор будет выполнять те дизъюнктивные члены дизъюнкции $D(Y)$, в которые входят пары литералов с теми же парами индексов, что и у множеств $\mathscr{A}_{i_1i_2}$ из $\mathscr{K}_\beta$. Взаимно-однозначное соответствие выполняющих наборов функции $\mathscr{D}(Y)$ и множеств, порождаемых пересечением множеств из Ω, установлено. Значит, $\#\{\alpha/D(\alpha)=1\}$ совпадает с числом множеств, порождаемых пересечением множеств из Ω, и число выполняющих наборов функции $\mathscr{F}(Y)$ равно $\#\{\eta/\mathscr{F}(\delta)=1\}=$ $=2^n-\#\{\alpha/D(\alpha)=1\}$. Тем самым задача по определению числа решений задачи «выполнимость монотонной 2-КНФ» сведена к задаче подсчета числа всевозможных пересечений семейства множеств, т. е. к задаче о числе ДСМ-гипотез. При сведении требуется построить r множеств $\mathscr{A}_{i_1i_2}$ и для каждого $\mathscr{A}_{i_1i_2}$ — построить не более $(n-2)$ множеств из $\mathscr{B}(\mathscr{A}_{i_1i_2})$. Значит, сводимость полиномиальна (порядка $O(r \cdot n^2)$) и задача «найти число всех гипотез» $\#P$-полна.

Следствие. Перечислительные задачи, соответствующие задачам 1—3 с функционалами h, l, $h+l$, $\#P$-полны.

В самом деле принадлежность указанных задач классу $\#P$ очевидна. В то же время, умея находить число решений задачи 3, мы можем решить задачу «найти число всех гипотез», просуммировав числа решений задачи 3 по k для $1 \leqslant k \leqslant |U^1|$ (для функционала h), для $1 \leqslant k \leqslant |\Omega|$ (для функционала l) и $1 \leqslant k \leqslant |U^1|+|\Omega|$ (для функционала $h+l$). Перечислительная задача, соответствующая задаче 2, является более общей задачей, чем задача «найти число всех гипотез» (которая получается из указанной, когда $k=1$).

Заметим, что при доказательстве теоремы 4 мы могли бы погрузить U^1 в множество U_*^1, большее по мощности в полиномиальное число раз, и либо добавить ко всем элементам из $\mathcal{B}(\mathcal{A}_{ij})$ по одинаковому подмножеству из U^1, либо не делать этого, не нарушив при этом полиномиальности сводимости. Принимая во внимание это соображение, можно сформулировать более сильную, чем теорема 4, теорему.

Теорема 5.

1. Задача «найти число всех гипотез» со значением функционала h, не превышающим $\alpha \cdot |U|^\beta$, $(0<\alpha<1, 0<\beta\leqslant 1)$, является $\#P$-полной.

2. Задача «найти число всех гипотез» со значением функционала h, не меньших $\alpha \cdot |U^1|^\beta$ $(0<\alpha<1, 0<\beta\leqslant 1)$, является $\#P$-полной.

Аналогичные теоремы могут быть сформулированы для функционалов l, $h+l$ (с заменой $|U^1|$ в формулировке на $|\Omega|$ и $|U^1|+|\Omega|$, соответственно).

* * *

В заключение автор хотел бы выразить признательность А. Ю. Когану и Д. П. Скворцову за ряд полезных указаний и внимание, проявленное к работе.

СПИСОК ЛИТЕРАТУРЫ

1. Финн В. К. Правдоподобные выводы и правдоподобные рассуждения // Итоги науки и техники. Сер. Теория вероятностей, математическая статистика, теоретическая кибернетика.— М.: ВИНИТИ, 1988.— Т. 28.— С. 1—86.
2. Аншаков О. М. Скворцов Д. П. Финн В. К. Логические средства экспертных систем типа ДСМ // Семиотика и информатика.— 1986.— Вып. 28.— С. 65—101.
3. Забежайло М. И. Финн В. К. и др. Об одном методе автоматического формирования гипотез и его программной реализации // НТИ. Сер. 2.— 1982.— № 4.— С. 20—26.
4. Михеенкова М. А. Финн В. К. Об одном классе экспертных систем с неполной информацией // Изв. АН СССР. Сер. Тех. киб.— 1986.— № 5.— С. 82—103.
5. Журавлев Ю. И. Алгоритмы построения минимальных дизъюнктивных форм для функции алгебры логики // Дискретная математика и математические вопросы кибернетики / Под общ. ред. С. В. Яблонского, О. Б. Лупанова.— М.: Наука, 1974.— С. 67—98.
6. Левит В. Е. Переверзев-Орлов В. С. Структура и поле данных при распознавании образов— М.: Наука, 1984.
7. Эрдёш П. Спенсер Дж. Вероятностные методы в комбинанорике— М.: Мир, 1976.
8. Забежайло М. И. О некоторых переборных задачах, возникающих при автоматическом порождении гипотез ДСМ-методом // НТИ Сер. 2.— 1988 — № 1.— С. 28—31.
9. Гэри М. Джонсон Д. Вычислительные машины и труднорешаемые задачи.— М.: Мир, 1982.— 416 с.
10. Харари Ф. Теория графов.— М.: Мир, 1973.— 302 с.
11. Левит В. Е. Алгоритм поиска подматрицы максимального периметра, состоящей из единиц на 0—1 матрице // Система передачи и обработки информации: Сб. тр. / ИППИ АН СССР.— М., 1988.— С. 42—45.
12. Пападимитриу Х. Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность — М.: Мир, 1985.— 512 с.
13. Karp R. Hopcroft G. A. 2.5-algorithm for maximum matching in bipartite graphs // SIAM J. Comput.— 1973.— Vol. 2, № 2.— P. 225—231.
14. Valiant L. G. The complexity of computing the permanent // Theoretical Comp. Sci.— 1979.— № 8.— P. 189—201.
15. Valiant L. G. The complexity of enumeration and reliability problems // SIAM J. Comput.— 1979.— Vol. 8, № 1.— 410—421.
16. Simon J. On the difference between one and many // Lect. Not. Comp. Sci.— 1977.— Vol. 52 — P. 480—491.

Материал поступил в редакцию 21.12.88.