Подсчет числа выполняющих наборов для ДНФ

Н. Н. Кузюрин С. А. Фомин

12 декабря 2010 г.

$$DNF = (x_3 \cdot \bar{x}_4 \cdot \bar{x}_1) \vee (x_1 \cdot x_4) \vee (x_4 \cdot \bar{x}_2) \vee (x_3 \cdot x_4 \cdot \bar{x}_2) \vee (x_3 \cdot \bar{x}_2)$$

x_1	x_2	x_3	x_4	c_1	c_2	c_3	c_4	c_5	ДНФ
0	0	0	0	·	·	·	·	·	·
1	0	0	0	·	·	·	·	·	·
0	1	0	0	·	·	·	·	·	·
1	1	0	0	·	·	·	·	·	·
0	0	1	0	■	·	·	·	□	1
1	0	1	0	·	·	·	·	■	1
0	1	1	0	■	·	·	·	·	1
1	1	1	0	·	·	·	·	·	·
0	0	0	1	·	·	■	·	·	1
1	0	0	1	·	■	□	·	·	1
0	1	0	1	·	·	·	·	·	·
1	1	0	1	·	■	·	·	·	1
0	0	1	1	·	·	■	□	□	1
1	0	1	1	·	■	□	□	□	1
0	1	1	1	·	·	·	·	·	·
1	1	1	1	·	■	·	·	·	1

Полиномиальная рандомизированная аппроксимационная схема подсчета числа выполняющих наборов.

Определение

Пусть

$$\begin{aligned} Z &\ \text{— перечислительная задача.} \\ I &\ \text{— вход для } Z,\ n = |I|. \\ \#(\mathrm{I}) &\ \text{— число различных решений для входа } I \text{ задачи } Z. \\ \varepsilon > 0 &\ \text{— параметр точности.} \end{aligned}$$

*Вероятностный алгоритм $A(I, \varepsilon)$ — **полиномиальная рандомизированная аппроксимационная схема**, если время его работы ограничено полиномом от n, и*

$$\mathsf{P}[(1-\varepsilon)\#(I) \le A(I,\varepsilon) \le (1+\varepsilon)\#(I)] \ge \frac{3}{4}.$$

Определение

Полностью полиномиальной рандомизированной аппроксимационной схемой *(FPRAS) называется полиномиальная рандомизированная аппроксимационная схема, время работы которой ограничено полиномом от n и $1/\varepsilon$.*

Определение

Полностью полиномиальной рандомизированной аппроксимационной схемой с параметрами (ε, δ) *(или кратко* (ε, δ)*-*$\mathcal{FPRAS}$*) для перечислительной задачи Z называется полностью полиномиальная рандомизированная аппроксимационная схема, которая на каждом входе I вычисляет* ε*-аппроксимацию для* $\#(I)$ *с вероятностью не менее* $1 - \delta$ *за время, полиномиальное от* n*,* $1/\varepsilon$ *и* $\log 1/\delta$*.*

Задача

$f(x_1, \ldots, x_n) = C_1 \vee \ldots \vee C_m$ — булева формула в дизъюнктивной нормальной форме (ДНФ), где каждая скобка C_i есть конъюнкция $L_1 \wedge \ldots \wedge L_{k_i}$ k_i литералов (см. определение).

*Набор $\mathbf{v} = (v_1, \ldots, v_n)$ — **выполняющий для** f, если $f(v_1, \ldots, v_n) = 1$.*

Найти число выполняющих наборов для данной ДНФ.

Стандартный метод Монте-Карло

G: множество выполняющих наборов.

V: множество всех двоичных наборов длины n

1. Проведем N независимых испытаний:
   - выбирается случайно $v \in V$;
   - $y_i = f(v_i)$.
2. Рассмотрим сумму независимых случайных величин

$$Y = \sum_{i=1}^{N} y_i$$

Аппроксимация $|G|$:

$$\tilde{G} = \frac{Y}{N}|V|$$

Монте-Карло — не (ε, δ)-$\mathcal{FPRAS}$

$$p = P\{y_i = 1\} = \frac{|G|}{|V|}$$

Лемма

Пусть $x_1, \ldots, x_n$ *— независимые случайные величины:*

$$P\{x_i = 1\} = p,$$
$$P\{x_i = 0\} = 1 - p.$$

Тогда

$$X = \sum_{i=1}^{n} x_i, \quad \mathbf{E}\, X = np.$$

А для $\forall\delta$, $0 < \delta < 1$, *выполнено:*

$$P\{|X - \mathbf{E}\, X| > \delta\, \mathbf{E}\, X\} \leq 2 \exp\{-(\delta^2/3)\, \mathbf{E}\, X\}.$$

Монте-Карло: Сколько испытаний сделать?

Оценим вероятность того, что аппроксимация хорошая:

$$\mathsf{P}\{(1-\varepsilon)|G| \le \tilde{G} \le (1+\varepsilon)|G|\} =$$

$$= \mathsf{P}\{(1-\varepsilon)\frac{N|G|}{|V|} \le Y \le (1+\varepsilon)\frac{N|G|}{|V|}\} =$$

$$= \mathsf{P}\{(1-\varepsilon)Np \le Y \le (1+\varepsilon)Np\}.$$

Применяя неравенства из леммы, получаем:

$$\mathsf{P}\{(1-\varepsilon)|G| \le \tilde{G} \le (1+\varepsilon)|G|\} > 1 - 2\exp\left\{-\frac{\varepsilon^2}{3}Np\right\}.$$

Потребуем, чтобы эта оценка была больше $1-\delta$:

$$2\exp\left\{-\frac{\varepsilon^2}{3}Np\right\} < \delta,$$

$$N > \frac{1}{p}\frac{3}{\varepsilon^2}\ln\frac{2}{\delta}.$$

Монте-Карло — не (ε, δ)-$\mathcal{FPRAS}$

Почему оценка не столь хороша?

$$N > \frac{1}{\mathbf{p}} \frac{3}{\varepsilon^2} \ln \frac{2}{\delta}.$$

p может быть экспоненциально мало (например 1 лишь в одной точке).

Обобщение

Пусть V — конечное множество, $H_1, \ldots, H_m \subseteq V$:

1. $\forall i$ $|H_i|$ вычислимо за полиномиальное время.
2. $\forall i$ возможно выбрать случайно и равномерно элемент из H_i.
3. $\forall v \in V$ за полиномиальное время проверяемо «$v \overset{?}{\in} H_i$».

Оценить $|H|$, где $H = H_1 \cup \ldots \cup H_m$.

Связь с исходной задачей: $H_i = \{v \in V : C_i(v) = 1\}$,

1. (1): r_i — число литералов в C_i: $\begin{cases} C_i \equiv 0 \Rightarrow |H_i| = 0 \\ C_i \not\equiv 0 \Rightarrow |H_i| = 2^{n-r_i} \end{cases}$
2. (2): зафиксировать переменные из C_i, а остальные — случайно.
3. (3): вычислить $C_i(v)$.

Модификация Монте-Карло: Идея

Универсум U образуется из точек H, *причем точка берется с кратностью, равной числу множеств H_i, которым она принадлежит*.

Тогда

$$|U| = \sum_{i=1}^{m} |H_i| \geq |H| \geq \frac{1}{m} \sum_{i=1}^{m} |H_i| = \frac{1}{m}|U|$$

К U можно применять стандартный алгоритм Монте-Карло.

$$DNF = (x_3 \cdot \bar{x}_4 \cdot \bar{x}_1) \vee (x_1 \cdot x_4) \vee (x_4 \cdot \bar{x}_2) \vee (x_3 \cdot x_4 \cdot \bar{x}_2) \vee (x_3 \cdot \bar{x}_2)$$

x_1	x_2	x_3	x_4	C_1	C_2	C_3	C_4	C_5	ДНФ
0	0	0	0	·	·	·	·	·	·
1	0	0	0	·	·	·	·	·	·
0	1	0	0	·	·	·	·	·	·
1	1	0	0	·	·	·	·	·	·
0	0	1	0	■	·	·	·	□	1
1	0	1	0	·	·	·	·	■	1
0	1	1	0	■	·	·	·	·	1
1	1	1	0	·	·	·	·	·	·
0	0	0	1	·	·	■	·	·	1
1	0	0	1	·	■	□	·	·	1
0	1	0	1	·	·	·	·	·	·
1	1	0	1	·	■	·	·	·	1
0	0	1	1	·	·	■	□	□	1
1	0	1	1	·	■	□	□	□	1
0	1	1	1	·	·	·	·	·	·
1	1	1	1	·	■	·	·	·	1

$$U = \{(v, i) | \; v \in H_i\}$$

$$|U| = \sum_{i=1}^{m} |H_i| \geq |H|$$

$$\operatorname{cov}(v) = \{(v, i) | \; (v, i) \in U\}$$

$$\operatorname{cov}(v) \leq m$$

$$U = \cup_{v \in H} \operatorname{cov}(v)$$

$$|U| = \sum_{v \in H} |\operatorname{cov}(v)|$$

$$G = \{(v, j_v \equiv \min_{(v,j) \in U} j)\}$$

$$|G| = |H|$$

(ε, δ)-$\mathcal{FPRAS}$

1. выбираем случайно и равномерно $u \in U$
2. $y_i = 1$ при $v \in G$, иначе $y_i = 0$.
3. $Y = \sum_{i=1}^{N} y_i$
4. $\tilde{G} = \frac{Y}{N}|U|$

Теорема

Метод Монте-Карло дает (ε, δ)-$\mathcal{FPRAS}$ для оценки $|G|$ при условии

$$N \geq \frac{3m}{\varepsilon^2} \ln \frac{2}{\delta}.$$

Время полиномиально по N.

Доказательство

Выбор случайного элемента (v, i) равномерно из U:

1. выберем i, $1 \le i \le m$ с вероятностью $\frac{|H_i|}{|U|} = \frac{|H_i|}{\sum_{i=1}^{m} |H_i|}$
2. выбирается случайно и равномерно $v \in H_i$.

Лемма

Получаемая пара (v, i) равномерно распределена в U.

Для доказательства достаточно перемножить соответствующие вероятности.

Проверка $(v, i) \overset{?}{\in} G$ — полиномиальна: $\forall j < i : C_j(v) = 0$.

Завершение доказательства

Из полученных ранее оценок о числе испытаний стандартного метода Монте-Карло имеем

$$N > \frac{1}{p}\frac{3}{\varepsilon^2}\ln\frac{2}{\delta}.$$

$$|U| = \sum_{v \in H} |\mathrm{cov}(v)| \le \sum_{v \in H} m \le m|H| = m|G|$$

Таким образом:

$$p = \frac{|G|}{|U|} \ge \frac{1}{m}$$

Получаем, что достаточное число испытаний:

$$N > \frac{3m}{\varepsilon^2}\ln\frac{2}{\delta}$$

Интернет поддержка курса

http://discopal.ispras.ru/

Вопросы?