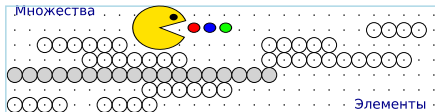


Точность жадного алгоритма для почти всех ВХОДОВ

Н. Н. Кузюрин С. А. Фомин

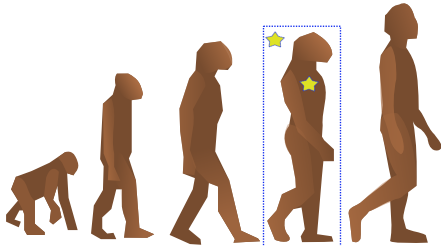
27 сентября 2021 г.

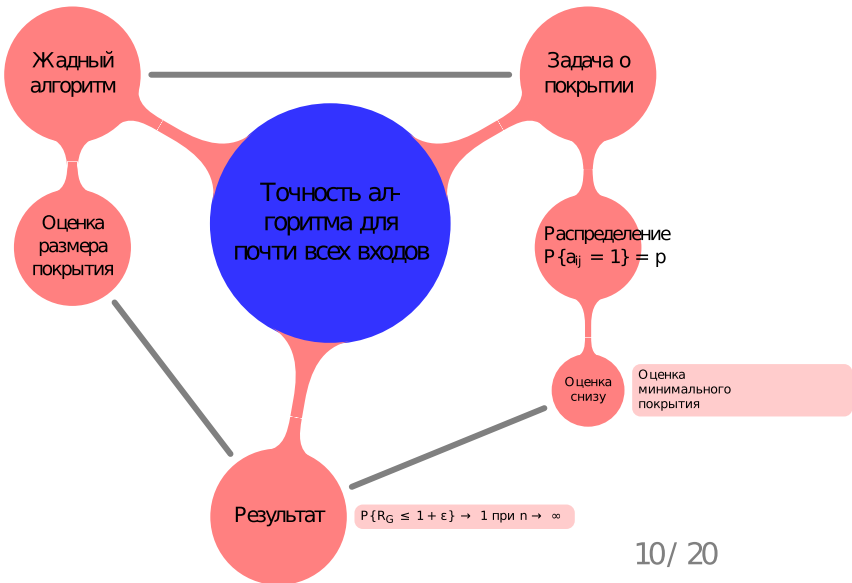


Приближенный алгоритм с гарантированной точностью

Определение

*Алгоритм называется **C-приближенным**, если при любых исходных данных он находит допустимое решение со значением целевой функции, отличающимся от оптимума не более чем в C раз.*





Покрытие множества

Задача

«Покрытие множества»^a.

- Множество X , $|X| = m$.
- Семейство подмножеств $\{S_1, \dots, S_n\}$, $S_j \subseteq X$.
- $X = \bigcup_{j=1}^n S_j$ (покрытие гарантировано).

Найти минимальное по мощности множество подмножеств, покрывающее X :

$$|J| \rightarrow \min,$$

$$J \subseteq \{1, 2, \dots, n\},$$

$$X = \bigcup_{j \in J} S_j.$$

^aВ англоязычной литературе — *Set Cover*.

Инцидентность

Определение

Пусть

$L = \{l_1, \dots, l_m\}$ — m -элементное множество;

$\{S_1, \dots, S_n\}$ — семейство подмножеств L ;

Тогда

- Элемент l_i и подмножество S_j **инцидентны**, если $l_i \in S_j$.
- **Матрицей инцидентности** называется $\{0,1\}$ -матрица $A = (a_{ij})$ размером $m \times n$, для которой:
 $a_{ij} = 1 \Leftrightarrow$ элемент l_i и подмножество S_j инцидентны.

ЦЛП для «Покрытия»

$$\sum_{j=1}^n x_j \rightarrow \min$$

$$\forall j: 1 \leq j \leq n \quad x_j \in \{0, 1\}$$

$$\forall i: 1 \leq i \leq m \quad \sum_{j=1}^n a_{ij} x_j \geq 1$$

ЦЛП для «Покрытия»

$$\sum_{j=1}^n x_j \rightarrow \min$$

$$\forall j: 1 \leq j \leq n \quad x_j \in \{0, 1\}$$

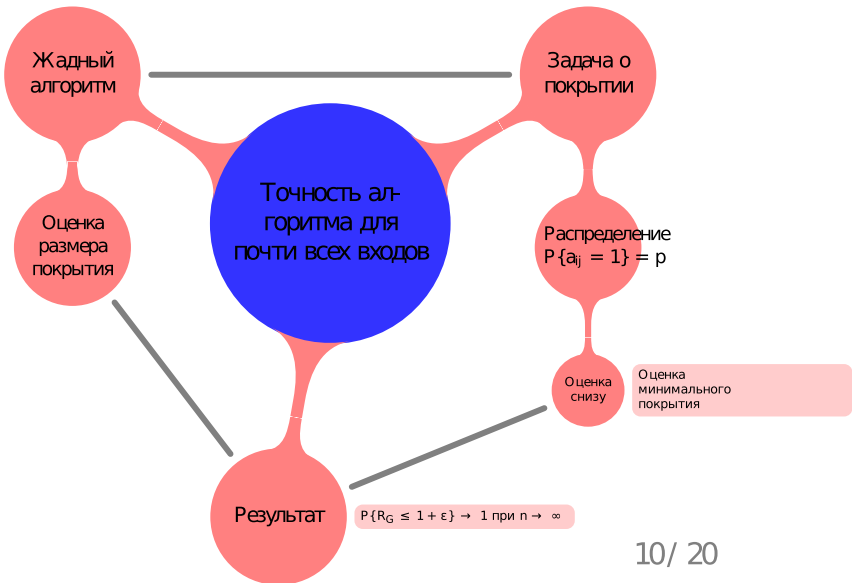
$$\forall i: 1 \leq i \leq m \quad \sum_{j=1}^n a_{ij}x_j \geq 1$$

ЦЛП для «Покрытия» (общий случай)

$$\sum_{j=1}^n c_j x_j \rightarrow \min$$

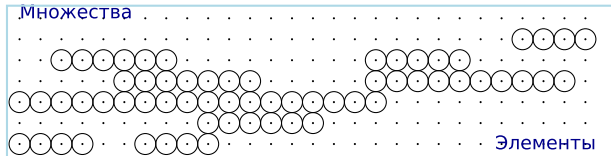
$$\forall j: 1 \leq j \leq n \quad x_j \in \{0, 1\}$$

$$\forall i: 1 \leq i \leq m \quad \sum_{j=1}^n a_{ij}x_j \geq b_i$$



10/ 20

Жадный алгоритм в задаче о покрытии



Выбор на
каждом шаге
подмножества,
покрывающего
максимальное
число еще
непокрытых
элементов:

Теорема

Пусть a_{ij} являются независимыми случайными величинами, принимающими значения $\{0, 1\}$, причем выполняется:

$$P\{a_{ij} = 1\} = p$$

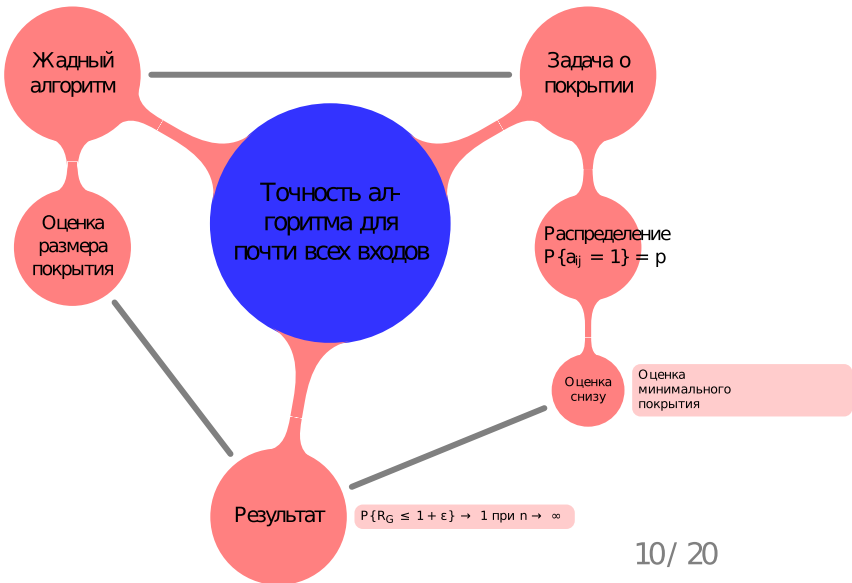
$$P\{a_{ij} = 0\} = 1 - p$$

$$\forall \gamma > 0 \quad \frac{\ln n}{m^\gamma} \rightarrow 0 \text{ при } n \rightarrow \infty$$

$$\frac{\ln m}{n} \rightarrow 0 \text{ при } n \rightarrow \infty$$

Тогда для любого фиксированного $\varepsilon > 0$

$$P\{R_G \leq 1 + \varepsilon\} \rightarrow 1 \text{ при } n \rightarrow \infty.$$



L : Множество столбцов.

Покрывают_ L : Множество столбцов L является покрытием.

Лемма

$$P(\text{Покрывают}_L) \leq \exp \left\{ -m(1 - p)^{|L|} \right\}$$

Доказательство.

Вероятность того, что столбцы из L :

- не покрывают какую-то одну строку:

L : Множество столбцов.

Покрывают L : Множество столбцов L является покрытием.

Лемма

$$P(\text{Покрывают } L) \leq \exp \left\{ -m(1 - p)^{|L|} \right\}$$

Доказательство.

Вероятность того, что столбцы из L :

- не покрывают какую-то одну строку: $(1 - p)^{|L|}$.
- покрывают какую-то одну строку:

L : Множество столбцов.

Покрывают L : Множество столбцов L является покрытием.

Лемма

$$P(\text{Покрывают } L) \leq \exp \left\{ -m(1 - p)^{|L|} \right\}$$

Доказательство.

Вероятность того, что столбцы из L :

- не покрывают какую-то одну строку: $(1 - p)^{|L|}$.
- покрывают какую-то одну строку: $1 - (1 - p)^{|L|}$.
- покрывают все m -строк:

L : Множество столбцов.

Покрывают L : Множество столбцов L является покрытием.

Лемма

$$P(\text{Покрывают } L) \leq \exp \left\{ -m(1 - p)^{|L|} \right\}$$

Доказательство.

Вероятность того, что столбцы из L :

- не покрывают какую-то одну строку: $(1 - p)^{|L|}$.
- покрывают какую-то одну строку: $1 - (1 - p)^{|L|}$.
- покрывают все m -строк: $(1 - (1 - p)^{|L|})^m$.

L : Множество столбцов.

Покрывают_ L : Множество столбцов L является покрытием.

Лемма

$$P(\text{Покрывают}_L) \leq \exp \left\{ -m(1 - p)^{|L|} \right\}$$

Доказательство.

Вероятность того, что столбцы из L :

- не покрывают какую-то одну строку: $(1 - p)^{|L|}$.
- покрывают какую-то одну строку: $1 - (1 - p)^{|L|}$.
- покрывают все m -строк: $(1 - (1 - p)^{|L|})^m$.

Откуда:

$$\left(1 - (1 - p)^{|L|}\right)^m \leq \exp \left\{ -m(1 - p)^{|L|} \right\}$$



Пусть $X(l)$ — число покрытий размера l , тогда

$$\begin{aligned}\mathbf{E} X(l) &= \sum_{L:|L|=l} P\{\text{Покрывают}_L\} \\ &= \binom{n}{l} (1 - (1 - p)^l)^m \\ &\leq n^l \exp\{-m(1 - p)^l\}\end{aligned}$$

Пусть для некоторой $0 < \delta < 1$, а $l_0 = -\lceil (1 - \delta) \frac{\ln m}{\ln(1 - p)} \rceil$, тогда

$$\begin{aligned}\ln \mathbf{E} X(l_0) &\leq l_0 \ln n - m(1 - p)^{l_0} \\ &\leq -\frac{\ln m \ln n}{\ln(1 - p)} - m \left\{ (1 - p)^{-\frac{(1 - \delta) \ln m}{\ln(1 - p)}} \right\} \\ &\leq -\frac{\ln m \ln n}{\ln(1 - p)} - m^\delta \xrightarrow{n \rightarrow \infty} -\infty\end{aligned}$$

Будем подразумевать пределы «при $n \rightarrow \infty$ ».

Имеем $\mathbf{E}X(l_0) \rightarrow 0$.

По неравенству Чебышева:

$$P\{X(l_0) \geq 1\} \leq \mathbf{E}X(l_0) \rightarrow 0$$

Значит,

$$P\{\text{Нет покрытий размера } l_0\} \rightarrow 1,$$

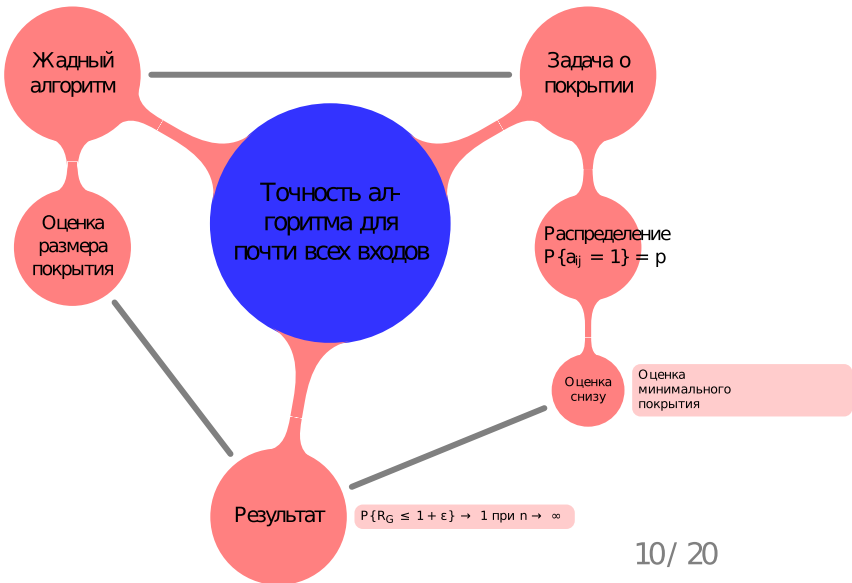
и

$$P\{M_{\text{optimal}} \geq l_0\} \rightarrow 1.$$

Получили:

Лемма

l_0 — нижняя вероятностная оценка размера минимального покрытия.



Лемма

Пусть Y — сумма n независимых случайных величин:

1 с вероятностью p ;

0 с вероятностью $1 - p$.

Тогда

$$P\{|Y - np| > \delta np\} \leq 2 \exp\left\{-\frac{\delta^2 np}{3}\right\}.$$

BadCol — событие «плохой столбец». (содержит $\leq (1 - \delta)pn$ единиц).

Тогда

$$P\{\text{BadCol}\} \leq 2 \exp\{-(\delta^2/3)np\}$$

X_{BadCols} — число плохих столбцов.

Тогда

$$\begin{aligned} \mathbf{E} X_{\text{BadCol}} &= m P\{\text{BadCol}\} \leq \\ &\leq 2 \exp\left\{\ln m - \frac{p\delta^2}{3}n\right\} = 2 \exp\{\ln m - \Omega(1)n\} \rightarrow 0 \end{aligned}$$

По неравенству Чебышева:

$$P\{X_{\text{BadCol}} \geq 1\} \leq \mathbf{E} X_{\text{BadCol}} \rightarrow 0$$

N_t — число непокрытых строк после t -го шага жадного алгоритма.

$$\begin{aligned} N_t &\leq N_{t-1} - \frac{N_{t-1}(1-\delta)pn}{n} = N_{t-1}(1 - (1-\delta)p) \leq \\ &\leq N_0(1 - (1-\delta)p)^t = m(1 - (1-\delta)p)^t. \end{aligned}$$

Максимальное t , при котором еще есть непокрытый элемент:

$$m(1 - (1-\delta)p)^t \geq 1$$

$$t \leq -\frac{\ln m}{\ln(1 - (1-\delta)p)}$$

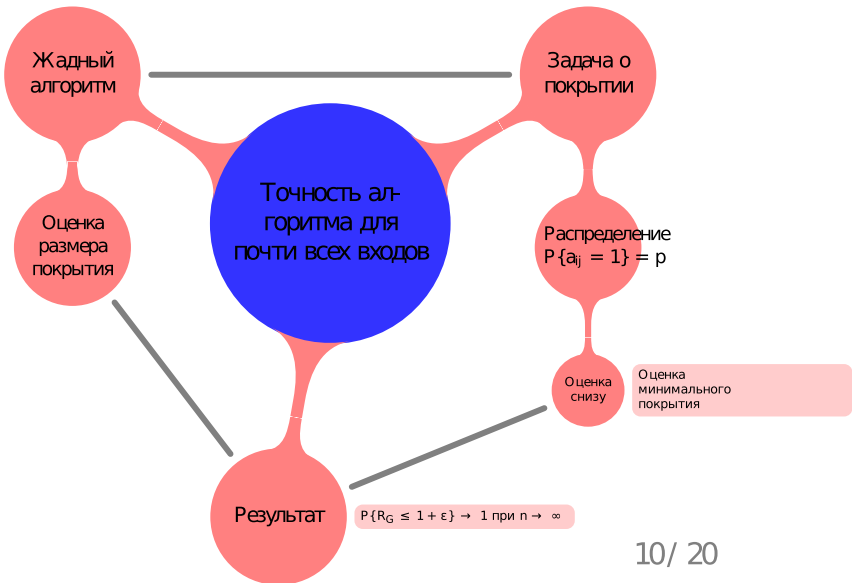
Лемма

$$P\left(Z_G \leq 1 - \frac{\ln m}{\ln(1 - p(1-\delta))}\right) \rightarrow 1.$$

$\forall \varepsilon > 0, P\{R_G \leq 1 + \varepsilon\} \rightarrow 1$ при $n \rightarrow \infty$

Оцениваем при $n \rightarrow \infty$:

$$\begin{aligned} R_G \leq \frac{Z_G}{l_0} &\leq \frac{1 - \frac{\ln m}{\ln(1-\rho(1-\delta))}}{-(1-\delta)\frac{\ln m}{\ln(1-\rho)}} \\ &\leq \frac{\ln(1-\rho)}{(1-\delta)\ln(1-\rho(1-\delta))} + o(1) \\ &\leq (1-\delta)^{-1} + o(1) \\ &\leq 1 + \varepsilon \end{aligned}$$



<http://discopal.ispras.ru/>

Вопросы?