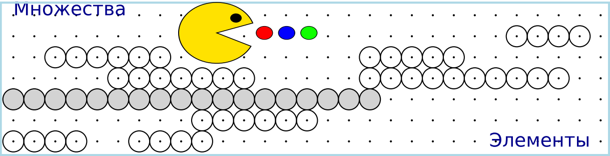
Точность жадного алгоритма для почти всех входов

Н. Н. Кузюрин С. А. Фомин

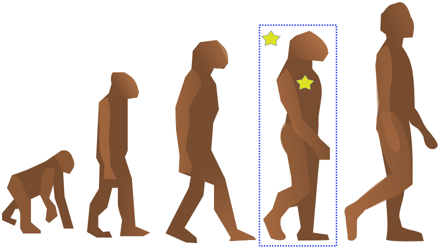
21 января 2024 г.

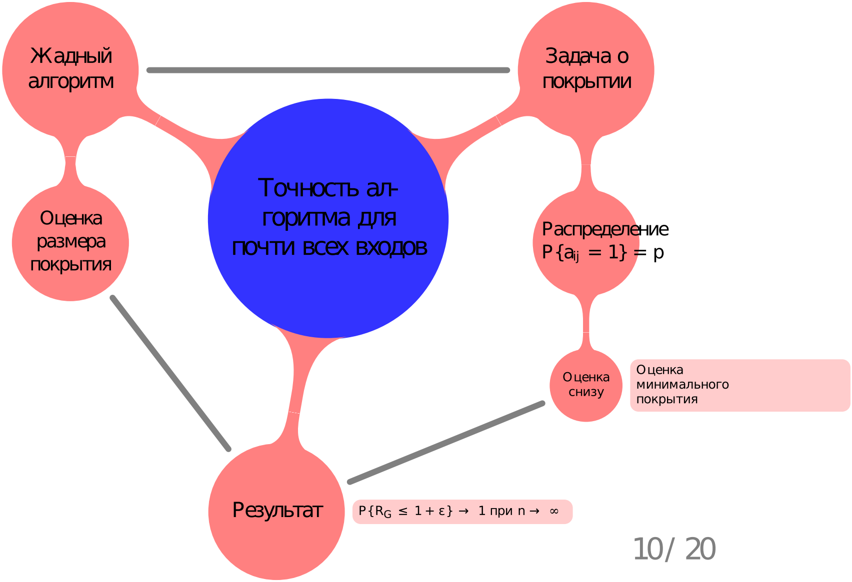


Приближенный алгоритм с гарантированной точностью

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.*




Жадный алгоритм
Задача о покрытии
Точность алгоритма для почти всех входов
Оценка размера покрытия
Распределение $P\{a_{ij} = 1\} = p$
Оценка снизу
Оценка минимального покрытия
Результат
$P\{R_G \le 1+\varepsilon\} \to 1$ при $n \to \infty$
10/ 20

Задача

«Покрытие множества»[a].

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

Найти минимальное по мощности множество подмножеств, покрывающее X:

$$
\begin{aligned}
&|J| \to \min,\\
&J \subseteq \{1, 2, \ \ldots, n\},\\
&X = \cup_{j \in J} S_j.
\end{aligned}
$$

[a] В англоязычной литературе — *Set Cover*.

Инцидентность

Определение

Пусть

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

$\{S_1, \ldots, S_n\}$ *— 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»

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

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

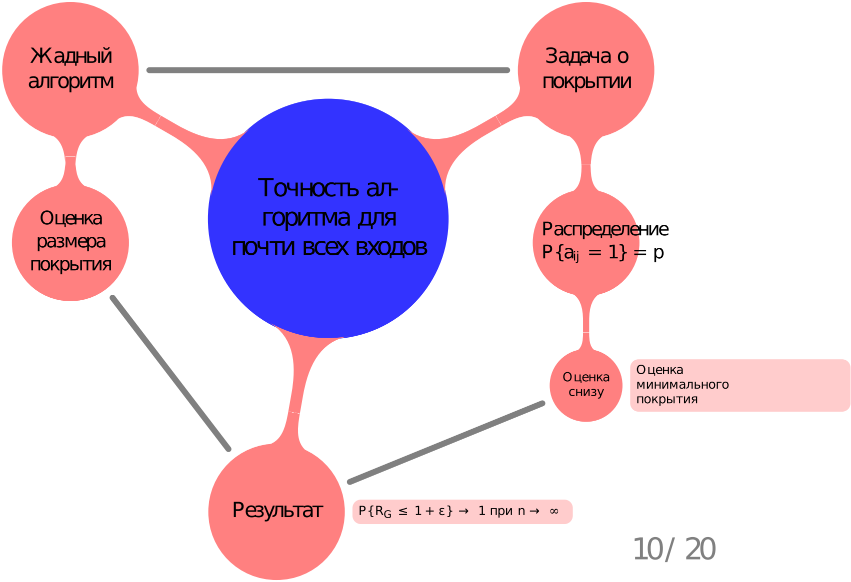
$$\forall i:\ 1 \le i \le m \quad \sum_{j=1}^{n} a_{ij} x_j \ge 1$$

ЦЛП для «Покрытия» (общий случай)

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

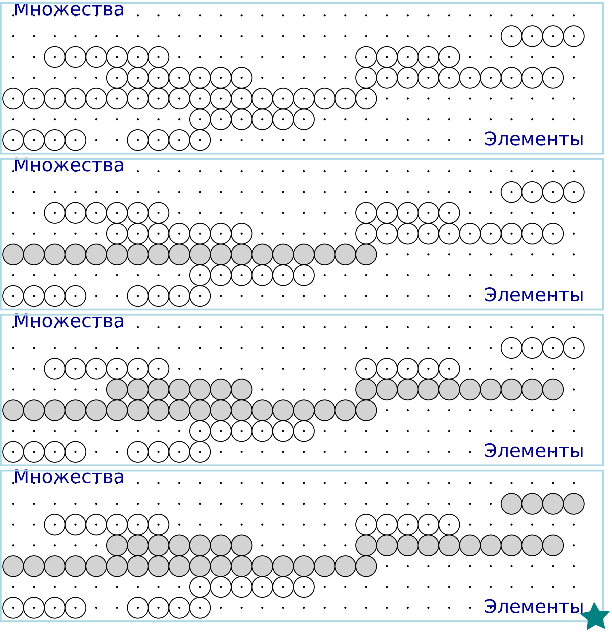
$$\forall j:\ 1 \le j \le n \quad x_j \in \{0,1\}$$

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



Жадный алгоритм
Задача о покрытии
Точность алгоритма для почти всех входов
Оценка размера покрытия
Распределение
$P\{a_{ij} = 1\} = p$
Оценка снизу
Оценка минимального покрытия
Результат
$P\{R_G \leq 1 + \varepsilon\} \to 1$ при $n \to \infty$
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:

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

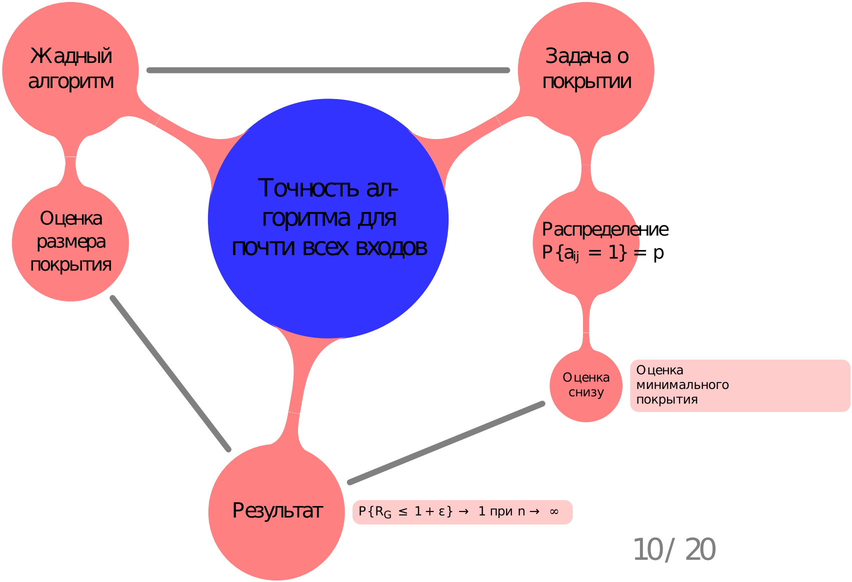
$$\mathsf{P}\{a_{ij} = 0\} = 1 - p$$

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

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

Тогда для любого фиксированного $\varepsilon > 0$

$$\mathsf{P}\{R_G \le 1 + \varepsilon\} \to 1 \text{ при } n \to \infty.$$


Жадный алгоритм
Задача о покрытии
Точность алгоритма для почти всех входов
Оценка размера покрытия
Распределение $P\{a_{ij} = 1\} = p$
Оценка снизу
Оценка минимального покрытия
Результат
$P\{R_G \leq 1 + \varepsilon\} \to 1$ при $n \to \infty$

L: Множество столбцов.

Покрывают_L: Множество столбцов L является покрытием.

Лемма

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

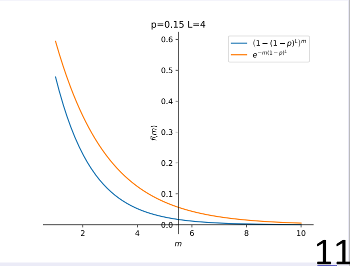
Доказательство.

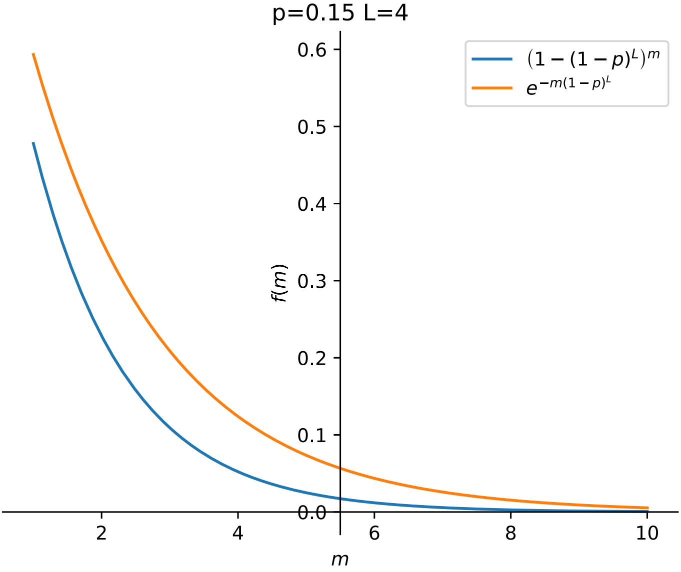
Вероятность того, что столбцы из L:

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

Откуда:

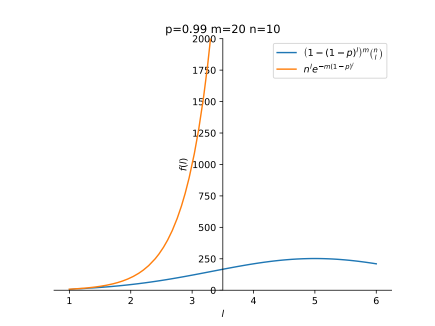
$$\left(1-(1-p)^{|L|}\right)^m \leq \exp\left\{-m(1-p)^{|L|}\right\}$$




p=0.15 L=4
$(1-(1-p)^L)^m$
$e^{-m(1-p)^L}$
f(m)
m
0.0
0.1
0.2
0.3
0.4
0.5
0.6
2
4
6
8
10

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

$$\mathbf{E}\, X(l) = \sum_{L:|L|=l} \mathrm{P}\{\text{Покрывают\_L}\}$$
$$= \binom{n}{l}\left(1-(1-p)^l\right)^m$$
$$\le n^l \exp\left\{-m(1-p)^l\right\}$$

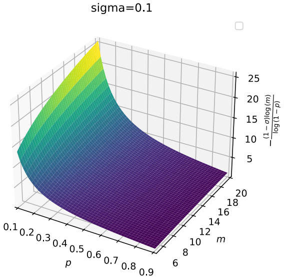


Грубоватая оценка сверху. Можно сделать лучше?

Пусть для некоторой $0 < \delta < 1$:

$$l_0 = -\left\lceil (1-\delta)\frac{\ln m}{\ln(1-p)} \right\rceil$$

тогда



$$\begin{aligned}
\ln \mathbf{E}\, X(l_0) &\le l_0 \ln n - m(1-p)^{l_0} \\
&\le -\frac{\ln m \ln n}{\ln(1-p)} - m\left\{(1-p)^{-\frac{(1-\delta)\ln m}{\ln(1-p)}}\right\} \\
&\le -\frac{\ln m \ln n}{\ln(1-p)} - m^{\delta} = -m^{\delta}\cdot\left(1 + \frac{1}{\ln(1-p)}\cdot\frac{\ln m}{m^{\frac{\delta}{2}}}\cdot\frac{\ln n}{m^{\frac{\delta}{2}}}\right) \\
&\underset{n\to\infty}{\longmapsto} -m^{\delta} \underset{n\to\infty}{\longmapsto} -\infty
\end{aligned}$$

Имеем $\mathbf{E}\, X(l_0) \longmapsto_{n\to\infty} 0$.

Будем подразумевать пределы «при $n \to \infty$». По неравенству Чебышева:

$$\mathsf{P}\{X(l_0) \geq 1\} \leq \mathbf{E}\, X(l_0) \to 0$$

Значит,

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

и

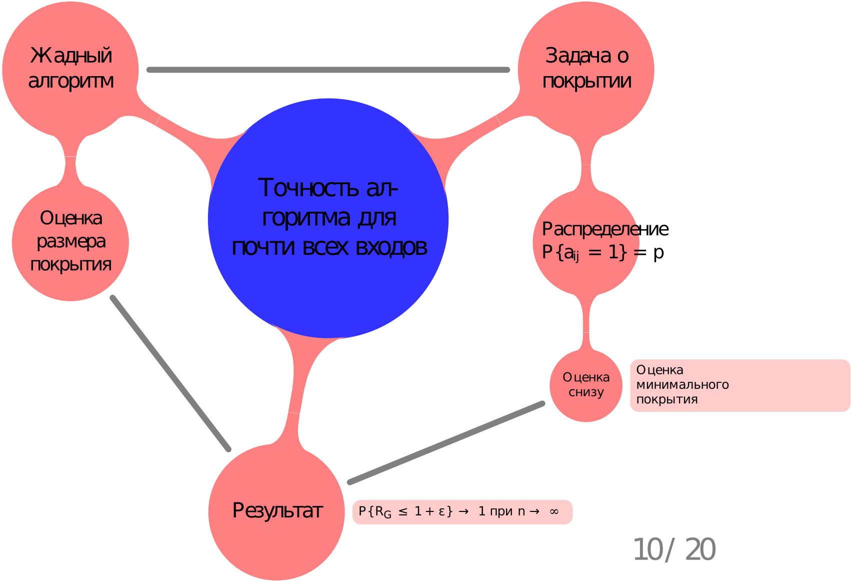
$$\mathsf{P}\{M_{optimal} \geq l_0\} \to 1.$$

Получили:

Лемма

l_0 — нижняя вероятностная оценка размера минимального покрытия.

Ура! Мы прошли по правой ветке карты, теперь пойдем по левой!


Жадный алгоритм
Задача о покрытии
Точность алгоритма для почти всех входов
Оценка размера покрытия
Распределение $P\{a_{ij} = 1\} = p$
Оценка снизу
Оценка минимального покрытия
Результат
$P\{R_G \le 1 + \varepsilon\} \to 1$ при $n \to \infty$

10/ 20

Будем использовать техническую лемму для биномиального распределения:

Лемма

Пусть Y — сумма n независимых случайных величин:

$$\begin{array}{ll} 1 & \text{с вероятностью } p; \\ 0 & \text{с вероятностью } 1 - p. \end{array}$$

Тогда

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

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

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

$X_{BadCols}$ — число плохих стобцов.

Тогда, в решающий момент используем заготовленные «условия неперекошенности»:

$$\mathbf{E}\, X_{BadCol} = m\,\mathsf{P}\{BadCol\} \leq$$
$$\leq 2\exp\left\{\ln m - \frac{p\delta^2}{3}n\right\} = 2\exp\{\ln m - \Omega(1)n\} \to 0$$

По неравенству Чебышева:

$$\mathsf{P}\{X_{BadCol} \geq 1\} \leq \mathbf{E}\, X_{BadCol} \to 0$$

N_t — число непокрытых строк после t-го шага жадного алгоритма. Для этих строк должен быть столбец с числом «1» не менее среднего:

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

Максимальное t, при котором еще есть непокрытый элемент:

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

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

И наконец, пройдя и по левой ветке, финишируем на следующем слайде.

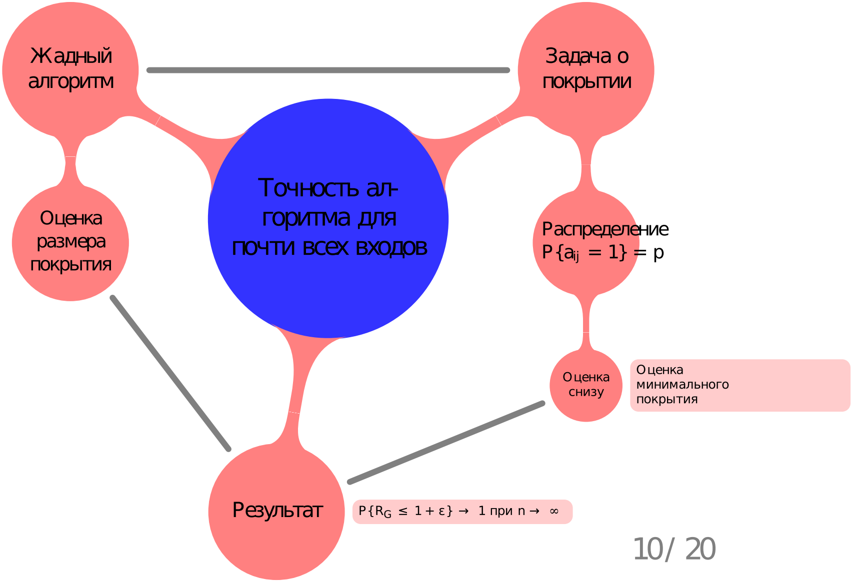
$\forall \varepsilon > 0$, $\mathrm{P}\{R_G \leq 1 + \varepsilon\} \to 1$ при $n \to \infty$

Лемма

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

Оцениваем при $n \to \infty$:

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


Жадный алгоритм
Задача о покрытии
Точность алгоритма для почти всех входов
Оценка размера покрытия
Распределение $P\{a_{ij} = 1\} = p$
Оценка снизу
Оценка минимального покрытия
Результат
$P\{R_G \le 1 + \varepsilon\} \to 1$ при $n \to \infty$
10/ 20

Интернет поддержка курса

http://discopal.ispras.ru/

Вопросы?