Лекция 11.

Оценки функциональной сложности
двумерной задачи о метрической близости
для метрики городских кварталов.

1 On Functional Complexity of Two-dimensional Manhattan Metrics Closeness Problem

We consider following complexity characteristics:

$$T'(U,x) = T(U,x) - |\mathcal{J}_I(x)|, \quad T'(U) = \mathbf{E}T'(U,x), \quad \widehat{T}'(U) = \max_{x\in X} T'(U,x).$$

$T'(U,x)$ characterizes the search time without the answer enumeration time for query x. It is not hard to prove that $T'(U) = T(U) - R(I)$.

In what follows we will write $\log k$ instead of $\log_2 k$.

We investigate the two-dimensional manhattan metrics closeness problem. Let $R \in [0,2]$, $X = [0,1]^2$ be the set of queries, $Y = [0,1]^2$ be the set of records, ρ be the relation on $X \times Y$ such that for $x = (x_1, x_2) \in X$ and $y = (y_1, y_2) \in Y$,

$$x\rho y \Longleftrightarrow |x_1 - y_1| + |x_2 - y_2| \le R.$$

The triple $S_m = \langle X, Y, \rho\rangle$ is called the type of two-dimensional manhattan metrics closeness problem.

Consider the following affinity of the plane:

$$\begin{cases} x'_1 &= (x_1 - x_2 + 1)/2 \\ x'_2 &= (x_1 + x_2)/2. \end{cases}$$

This affinity realizes the contraction in $\sqrt{2}$ times, the rotation by the angle $\pi/4$ about the point $(0,0)$, and the shift along abscissa by $1/2$.

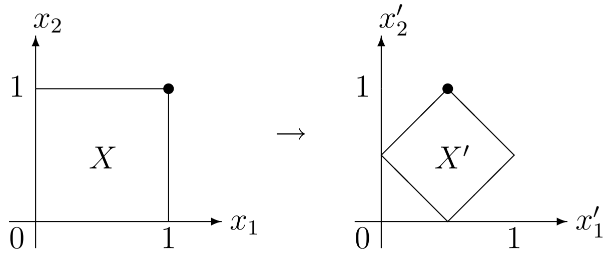

Рис. 1: Affinity of X to X'

This affinity maps the set X to

$$X' = \{(x_1, x_2) : (x_1 \in [0, \tfrac{1}{2}] \& \tfrac{1}{2} - x_1 \le x_2 \le \tfrac{1}{2} + x_1) \vee$$
$$\vee (x_1 \in [\tfrac{1}{2}, 1] \& x_1 - \tfrac{1}{2} \le x_2 \le \tfrac{3}{2} - x_1)\}$$

(see Figure 1), and the relation ρ to the relation ρ':

$$(x_1, x_2)\rho'(y_1, y_2) \iff |x_1 - y_1| \le r \& |x_2 - y_2| \le r,$$

where $r = R/2$.

It is clear that the type $S_m = \langle X, Y, \rho \rangle$ of the two-dimensional manhattan metrics closeness problem is equivalent to the type $S'_m = \langle X', X', \rho' \rangle$. To simplify the exposition we replace the relation ρ' by the relation ρ'' which is given by

$$(x_1, x_2)\rho''(y_1, y_2) \iff -r \le y_1 - x_1 < r \& |x_2 - y_2| \le r.$$

It is not hard to see that this replacement do not change the complexity of algorithms. Therefore we will research the type $S''_m = \langle X', X', \rho'' \rangle$.

Suppose the probability measure $\mathbf{P}$ is specified by a probability density $p(u, v)$. Let us extend $p(u, v)$ on $[0, 1]^2$ such that $p(u, v) = 0$ on $[0, 1]^2 \setminus X'$. Denote

$$p_{b,d}(v) = \int_b^d p(u, v)du \Big/ \int_b^d du \int_0^1 p(u, v)dv,$$

$$p_{b,d}(u, v) = p(u, v) \Big/ \int_b^d du \int_0^1 p(u, v)dv.$$

We say the probability density $p(u, v)$ is c-bounded if for any $b, d \in [0, 1]$ such that $b < d$, the following conditions hold:

- $(d-b)p_{b,d}(v) \le c$ for any $v \in [0,1]$;
- $(d-b)p_{b,d}(u,v) \le c$ for any $(u,v) \in [b,d] \times [0,1]$.

We write $g(n) = O(f(n))$ as $n \to \infty$ if there exist the constants C_1, C_2, and N_0 such that $f(n) \le C_1 g(n)$ and $g(n) \le C_2 f(n)$ for any $n \ge N_0$.

Let $\tilde{x} = (u,v) \in X'$,

$$G_1 = \left\{ g^{1,m}_{a,b}(\tilde{x}) = \left[\frac{u-a}{b-a} \cdot m \right] + 1,\ m \in \mathbb{N},\ 0 \le a < b \le 1 \right\}, \quad (1)$$

$$G_2 = \left\{ g^2_a(\tilde{x}) = \begin{cases} 1, & \text{if } u < a, \\ 2, & \text{if } u \ge a, \end{cases} \ a \in [0,1] \right\}, \quad (2)$$

$$G_3 = \left\{ g^3_a(\tilde{x}) = \begin{cases} 1, & \text{if } v \le a \\ 2, & \text{if } v > a \end{cases}, a \in [0,1] \right\}, \quad (3)$$

$$F = \left\{ f_a(\tilde{x}) = \begin{cases} 1, & \text{if } v - r \le a \le v + r \\ 0, & \text{otherwise} \end{cases}, a \in [0,1] \right\}. \quad (4)$$

$$f_{id}(\tilde{x}) \equiv 1, \quad (5)$$

$$\mathcal{F} = \langle F \cup \{f_{id}\}, G_1 \cup G_2 \cup G_3 \rangle. \quad (6)$$

The following assertion is valid.

Теорема 1. *Suppose the probability measure* $\mathbf{P}$ *is specified by a probability density* $p(u,v)$ *and* $p(u,v)$ *is* c*-bounded,* $I = \langle X', V, \rho'' \rangle \in S''_m$, $|V| = k$, $\mathcal{F}$ *is the base set given by (1)–(6); then for any natural* n *such that* $1 \le n \le \ln k$, *there exists an IG* $U \in \mathcal{U}(I, \mathcal{F})$ *such that* $T'(U) \le 7n-2$, $\widehat{T}'(U) = O(n \log k)$, $Q(U) = O(nk^{1+1/n})$ *as* $k \to \infty$.

2 Idea of theorem proof

The paper size limitation do not allow us to give a complete and exact proof of the theorem. Therefore we only describe the parametric algorithm, which gives the upper bound. We denote this algorithm by symbol $\mathcal{A}_n$, where n is the parameter of the algorithm.

If M is a set, then $|M|$ is equal to the number of elements of the set M.

If r is a real number, then let $[r]$ be the maximal integer no greater than r.

Let

$$V = \{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\} \quad (7)$$

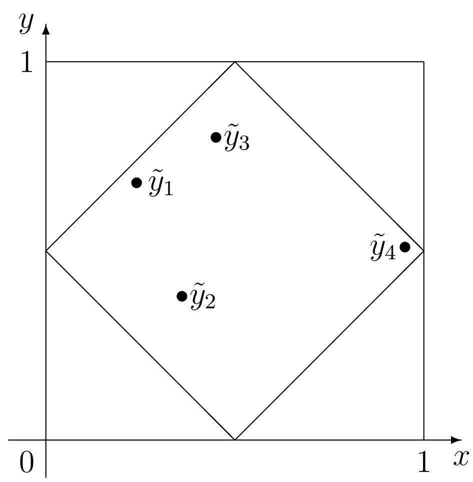

Рис. 2: Example of library

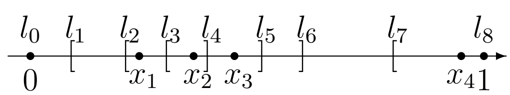

Рис. 3: Example of set $\mathcal{N}$

be a library. We can assume without loss of generality that $x_1 < x_2 < \dots < x_k$.

First let us describe the algorithm $\mathcal{A}_1$.

Denote

$$\mathcal{N} = \{0, 1, x_1 - r, x_1 + r, \dots, x_k - r, x_k + r\} \cap [0, 1]. \tag{8}$$

Let

$$\mathcal{N} = \{l_0, l_1, \dots, l_m\}, \tag{9}$$

where $l_0 < l_1 < \dots < l_m$. It is clear that $m \le 2k + 1$, $l_0 = 0$, $l_m = 1$.

Figure 2 shows an example of a library, where $\tilde{y}_i = (x_i, y_i)$, $i = 1, 2, 3, 4$. Then Figure 3 shows the set $\mathcal{N} = \{l_0, l_1, \dots, l_8\}$ for this library, where $l_0 = 0$, $l_1 = x_1 - r$, $l_2 = x_2 - r$, $l_3 = x_3 - r$, $l_4 = x_1 + r$, $l_5 = x_2 + r$, $l_6 = x_3 + r$, $l_7 = x_4 - r$, $l_8 = 1$.

Denote

$$V_i = \big\{(x_j, y_j) \in V : [l_{i-1}, l_i) \subset [x_j - r, x_j + r)\big\}, \tag{10}$$

where $i = 1, \dots, m$.

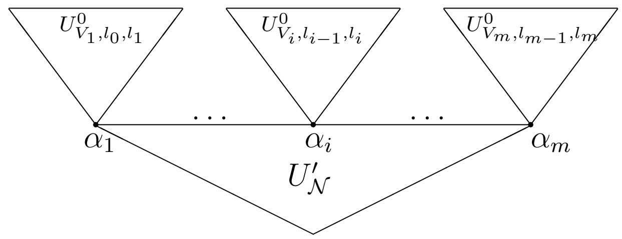

Рис. 4: IG $U^1_{V,\mathcal{N}}$

For our example from Figure 2, $V_1 = \emptyset$, $V_2 = \{\tilde{y}_1\}$, $V_3 = \{\tilde{y}_1, \tilde{y}_2\}$, $V_4 = \{\tilde{y}_1, \tilde{y}_2, \tilde{y}_3\}$, $V_5 = \{\tilde{y}_2, \tilde{y}_3\}$, $V_6 = \{\tilde{y}_3\}$, $V_7 = \emptyset$, $V_8 = \{\tilde{y}_4\}$.

It is easy to see that if $(u, v) \in X'$ and $u \in [l_i, l_{i+1})$, then $\neg((u, v)\rho''\tilde{y})$ for any $\tilde{y} \in V \setminus V_i$, i.e. if $u \in [l_i, l_{i+1})$, then the answer to the query (u, v) is a subset of V_i.

We sort all sets V_i $(i = 1, 2, \dots, m)$ in increasing order of the second coordinate (i.e. the y-coordinate).

Then for a query $(u, v) \in X'$, the algorithm $\mathcal{A}_1$ comprises from following steps.

1) We find in the set $\mathcal{N}$ minimal l_i such that $u \in [l_{i-1}, l_i)$. It can be done in 2 tacts in average and $O(\log m)$ tacts in the worst case.

2) We find in the set $V_i = \{(x'_1, y'_1), \dots (x'_q, y'_q)\}$ $(y'_1 < y'_2 < \dots < y'_q)$ minimal j such that $v - r \leq y'_j$. It can be done in 2 tacts in average and $O(\log q)$ tacts in the worst case.

3) While $y'_j < v + r$ we add the record (x'_j, y'_j) to the answer of the algorithm and set j=j+1. The time of the last step is equal to the time of the answer enumeration.

Describe this algorithm as an information graph.

Let $Z = \{z_0, z_1, \dots, z_q\} \subseteq [0, 1]$, where $z_0 < z_1 < \dots < z_q$. Denote by U'_Z the IG with leaves $\alpha_1, \dots, \alpha_q$ such that z_i is associated with the leaf α_i $(i = 1, 2, \dots, q)$ and

$$\varphi_{\alpha_i}(u, v) = \begin{cases} 1, & \text{if } z_{i-1} \leq u < z_i \\ 0, & \text{otherwise} \end{cases}.$$

We will construct the IG U'_Z by the method of binary searching with hash function. The IG U'_Z allows for query (u, v) to find minimal z_i such that $u < z_i$.

It follows that

$$Q(U'_Z) = (2+c)q + 1,\ T(U'_Z) \le 2,\ T(U'_Z, (u,v)) \le 2 + \log q \qquad (11)$$

for any query $(u,v) \in [z_0, z_q) \times [0,1]$.

Let $[b,d) \subseteq [0,1)$, $V' = \{(x_1,y_1),\dots,(x_q,y_q)\}$, where $x_i \in [b,d)$ for any $i \in \{1,2,\dots,q\}$ and $0 \le y_1 < y_2 < \dots < y_q \le 1$. Denote by $U''_{V',b,d}$ the IG with leaves $\beta_1,\dots,\beta_q$ such that (x_i,y_i) is associated with the leaf β_i $(i = 1,2,\dots,m)$ and

$$\varphi_{\beta_i}(u,v) = \begin{cases} 1, & \text{if } y_{i-1} < v - r \le y_i \le v + r \\ 0, & \text{otherwise} \end{cases},$$

where $y_0 = 0$. If we will construct the IG $U''_{V',b,d}$ the same way as the IG U'_Z and add the check $y_i \le v + r$, then

$$Q(U''_{V',b,d}) = (3+c)q + 1,\ T(U''_{V',b,d}) \le 3,\ T(U''_{V',b,d}, (u,v)) \le 3 + \log q \quad (12)$$

for any query $(u,v) \in [b,d) \times [0,1]$.

In the IG $U''_{V',b,d}$ for each $i \in \{1,2,\dots,q-1\}$, we add the edge (β_i, β_{i+1}) and associate the predicate $f_{y_{i+1}}(u,v)$ with this edge. The obtained IG is denoted by $U^0_{V',b,d}$. For any query $(u,v) \in [b,d) \times [0,1]$

$$\mathcal{J}_{U^0_{V',b,d}}(u,v) = \{(x_i,y_i) \in V' : y_i \in [v-r, v+r]\}.$$

Moreover,

$$Q(U^0_{V',b,d}) = (4+c)q,\ T(U^0_{V',b,d}) \le 3,\ \widehat{T}(U^0_{V',b,d}) \le 3 + \log q. \quad (13)$$

Let the library V be given by (7), the libraries V_i be given by (10), the set $\mathcal{N}$ be given by (8), (9). Consider the IG $U'_{\mathcal{N}}$. Let us remember that its leaves is denoted by $\alpha_1,\dots,\alpha_m$. For each $i \in \{1,2,\dots,m\}$, we construct the IG $U^0_{V_i,l_{i-1},l_i}$ and identify the root of $U^0_{V_i,l_{i-1},l_i}$ with the vertex α_i, i.e. now the IG $U^0_{V_i,l_{i-1},l_i}$ is growing from α_i and α_i is not a pole of the resulting IG. We denote the resulting IG by $U^1_{V,\mathcal{N}}$. The IG $U^1_{V,\mathcal{N}}$ is shown in Figure 4.

The IG $U^1_{V,\mathcal{N}}$ describes the algorithm $\mathcal{A}_1$.

Denote $d = (5+c)$. Without loss of generality it can be assumed that $m \ge k$. Using (11)–(13), it is easy to get that

$$Q(U^1_{V,\mathcal{N}}) \le (2+c)m + 1 + m(4+c)k \le dm^2,\ T'(U^1_{V,\mathcal{N}}) \le 5,$$
$$\widehat{T}'(U^1_{V,\mathcal{N}}) \le \log m + \log k + 5 \le 2\log m + 5 \qquad (14)$$

for any query $(u, v) \in X'$, i.e. the memory size of the algorithm $\mathcal{A}_1$ is not greater than $O(k^2)$, the average search time is not greater than 5 the search time in the worst case is equal to $O(\log k)$ (the search time is considered without the answer enumeration time).

These evaluations is matched to the results of Theorem 1 for parameter $n = 1$.

In what follows, we assume that if some real number is used as an natural, then this natural number is the closest natural to this real. For example, we consider $\sqrt{8}$ as 3.

Make the induction hypothesis: for any natural number $q < n$, we can construct the algorithm $\mathcal{A}_q$ represented by the IG $U^q_{V,\mathcal{N}}$ and

$$\begin{aligned}
&Q(U^q_{V,\mathcal{N}}) \leq dqm^{1+1/q},\ T'(U^q_{V,\mathcal{N}}) \leq 7q - 2, \\
&\widehat{T}'(U^q_{V,\mathcal{N}}) \leq \frac{3+q}{2} \log m + 7q.
\end{aligned} \tag{15}$$

The inequalities (14) prove that the induction hypothesis is valid for $q = 1$.

Let $q > 1$. Denote

$$\begin{aligned}
&p(q, j) = j \cdot m^{1-1/q},\ j = 0, 1, \ldots, m^{1/q}, \\
&\mathcal{N}^q = \{l_{p(q,j)}, j = 0, 1, \ldots, m^{1/q}\}, \\
&\mathcal{N}^q_s = \{l_j \in \mathcal{N} : j \in [p(q, s-1), p(q, s)]\}, \\
&V^q_s = \bigcap_{j=p(q,s-1)+1}^{p(q,s)} V_j,\ W^q_s = \Big(\bigcup_{j=p(q,s-1)+1}^{p(q,s)} V_j\Big) \setminus V^q_s,
\end{aligned}$$

$s = 1, 2, \ldots, m^{1/q}$.

For example, if $\mathcal{N} = \{l_0, l_1, \ldots, l_8\}$, then

$$\begin{aligned}
&\mathcal{N}^2 = \{l_0, l_3, l_6, l_8\},\ \mathcal{N}^2_1 = \{l_0, l_1, l_2, l_3\}, \\
&\mathcal{N}^2_2 = \{l_3, l_4, l_5, l_6\},\ \mathcal{N}^2_3 = \{l_6, l_7, l_8\}.
\end{aligned}$$

If V is the library described in Figure 2, then $V^2_1 = \emptyset$, $W^2_1 = \{\tilde{y}_1, \tilde{y}_2\}$, $V^2_2 = \{\tilde{y}_3\}$, $W^2_2 = \{\tilde{y}_1, \tilde{y}_2\}$, $V^2_3 = \emptyset$, $W^2_3 = \{\tilde{y}_4\}$.

It is easy to see that $\big|\big(V_j \cup V_{j+1}\big) \setminus \big(V_j \cap V_{j+1}\big)\big| = 1$ for any $j = \{1, 2, \ldots, m-1\}$. Therefore

$$\big|W^q_s\big| \leq p(q, s) - p(q, s-1) \leq m^{1-1/q} \tag{16}$$

for any $s = 1, 2, \ldots, m^{1/q}$.

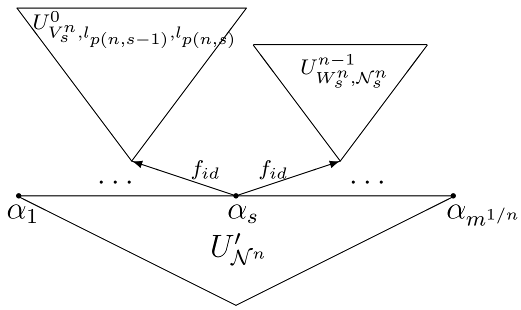

Рис. 5: IG $U^n_{V,\mathcal{N}}$

For a query $(u,v) \in X'$, the algorithm $\mathcal{A}_n$ comprises from following steps.

1) We find in the set $\mathcal{N}^n$ minimal s such that $u \in [l_{p(n,s-1}, l_{p(n,s)})$. It can be done in 2 tacts in average and $O(\log m^{1/n})$ tacts in the worst case.

2) We find in the set $V^n_s = \{(x'_1, y'_1), \dots (x'_t, y'_t)\}$ $(y'_1 < y'_2 < \dots < y'_t)$ the minimal number j such that $v - r \leq y'_j$. It can be done in 2 tacts in average and $O(\log t)$ tacts in the worst case. While $y'_j < v + r$ we add the record (x'_j, y'_j) to the answer of the algorithm and set j=j+1.

3) We apply the algorithm $\mathcal{A}_{n-1}$ for the query x and the sets $\mathcal{N}^n_s$ and W^n_s.

In Figure 5, the algorithm $\mathcal{A}_n$ is represented as the IG $U^n_{V,\mathcal{N}}$.

Calculate the complexity of the IG $U^n_{V,\mathcal{N}}$.

Using (15) and (16), we get

$$\begin{aligned}
Q(U^n_{V,\mathcal{N}}) &= \sum_{s=1}^{m^{1/n}} \big(2 + Q(U^0_{V^n_s, l_{p(n,s-1)}, l_{p(n,s)}}) + +Q(U^{n-1}_{W^n_s, \mathcal{N}^n_s})\big) + Q(U'_{\mathcal{N}^n}) = \\
&= (2+c)m^{1/n} + 1 + m^{1/n}\big((4+c)m + \\
&\quad +d(n-1)m^{(1-1/q)(1+1/(n-1))} + 2\big) \leq dnm^{1+1/n}, \\
T'(U^n_{V,\mathcal{N}}) &= 2 + 2 + 3 + 7(n-1) - 2 = 7n - 2, \\
\widehat{T}'(U^n_{V,\mathcal{N}}) &= \log m^{1/n} + 2 + 2 + \log k + 3 + \\
&\quad + \frac{3+n-1}{2} \log m^{1-1/n} + 7(n-1) = \\
&= \frac{3+n-1}{2} \log m + 7n.
\end{aligned}$$

This completes the proof of the theorem.