

# Результаты тестирования производительности ОАС-алгоритма трикластеризации (на штрих-операторах)

Реализация: Python 2.7.3

Машина: Core i3-370M, 2 x 2.4 ГГц, 3.87 Гб ОЗУ

ОС: Windows 7 x64

| Контекст | G   | M   | B    | # троек | Плотность |
|----------|-----|-----|------|---------|-----------|
| 1        | 30  | 30  | 30   | 2677    | 0.0991    |
| 2        | 250 | 795 | 22   | 3815    | 0.00087   |
| 3        | 51  | 924 | 2844 | 3037    | 2.266e-05 |

1) Контекст 1 (input\context\_gen\_30x30x30.txt)

| Порог плотности $\rho_{min}$ | T, сек – время исполнения алгоритма | N, шт – кол-во порожденных трикластеров |
|------------------------------|-------------------------------------|---|
| 0                            | 2.932                               | 2677                                    |
| 0.1                          | 2.917                               | 2677                                    |
| 0.2                          | 2.919                               | 2572                                    |
| 0.3                          | 2.936                               | 789                                     |
| 0.4                          | 2.919                               | 96                                      |
| 0.5                          | 2.924                               | 15                                      |
| 0.6                          | 2.889                               | 10                                      |
| 0.7                          | 2.92                                | 3                                       |
| 0.8                          | 2.91                                | 1                                       |
| 0.9                          | 2.934                               | 0                                       |
| 1                            | 2.896                               | 0                                       |

2) Контекст 2 (input\context\_gen\_250x795x22.txt)

| Порог плотности $\rho_{min}$ | T, сек – время исполнения алгоритма | N, шт – кол-во порожденных трикластеров |
|------------------------------|-------------------------------------|---|
| 0                            | 5.099                               | 3679                                    |
| 0.1                          | 5.145                               | 3679                                    |
| 0.2                          | 5.115                               | 3601                                    |
| 0.3                          | 5.154                               | 3227                                    |
| 0.4                          | 5.186                               | 2399                                    |
| 0.5                          | 5.165                               | 1438                                    |
| 0.6                          | 5.136                               | 654                                     |
| 0.7                          | 5.142                               | 297                                     |
| 0.8                          | 5.126                               | 244                                     |
| 0.9                          | 5.197                               | 234                                     |
| 1                            | 5.132                               | 234                                     |

### 3) Контекст 3 (input\context\_gen\_51x924x2844.txt)

| Порог плотности $\rho_{min}$ | T, сек – время исполнения алгоритма | N, шт – кол-во порожденных трикластеров |
|------------------------------|-------------------------------------|---|
| 0                            | 3.303                               | 2459                                    |
| 0.1                          | 3.266                               | 2459                                    |
| 0.2                          | 3.275                               | 2456                                    |
| 0.3                          | 3.271                               | 2449                                    |
| 0.4                          | 3.273                               | 2413                                    |
| 0.5                          | 3.274                               | 2305                                    |
| 0.6                          | 3.272                               | 2011                                    |
| 0.7                          | 3.268                               | 1287                                    |
| 0.8                          | 3.275                               | 1023                                    |
| 0.9                          | 3.299                               | 1001                                    |
| 1                            | 3.292                               | 1001                                    |

## Отличие в реализации по сравнению с описанием в статье «Экспериментальное сравнение алгоритмов трикластеризации»

### 1) В статье (алгоритм 1)

**Вход:**  $K = (G; M; B; I)$  \_ триконтекст;

$\rho_{min}$  - порог плотности

**Выход:**  $TSet = \{(X; Y; Z)\}$

1: **для всех**  $(g, m): g \in G, m \in M$

2:      $PrOA[g, m] = (g, m)'$

3: **для всех**  $(g, b): g \in G, b \in B$

4:      $PrOC[g, b] = (g, b)'$

5: **для всех**  $(m, b): m \in M, b \in B$

6:      $PrAC[m, b] = (m, b)'$

7: **для всех**  $(g, m, b) \in I$

8:      $T = (PrAC[m, b]; PrOC[g, b]; PrOA[g, m])$

9:      $Tkey = hash(T)$

10:    **если**  $Tkey \notin Tset.keys \wedge \rho(T) > \rho_{min}$  **то**

11:      $Tset[Tkey] = T$

### 2) Моя реализация (алгоритм 2)

**Вход:**  $K = (G; M; B; I)$  \_ триконтекст;

$\rho_{min}$  - порог плотности

**Выход:**  $TSet = \{(X; Y; Z)\}$

1: **для всех**  $(g, m, b) \in I$

2:      $T = ((g, m)', (g, b)', (m, b)')$

3:      $Tkey = hash(T)$

4:     **если**  $Tkey \notin Tset.keys \wedge \rho(T) > \rho_{min}$  **то**

5:      $Tset[Tkey] = T$

Я не понял, зачем в стр 1-6 алгоритма 1 считать много лишнего, если затем используются только значения (g,m,b) из I. Например, для контекста 3 ( $|G| = 51$ ,  $|M| = 924$ ,  $|B| = 2844$ ,  $|I| = 3037$ ), посчитаются  $|G||M| + |G||B| + |M||B| = 2820024$  значений массивов PrOA, PrOC и PrAC (в сумме), но потом, начиная со строки 7, используются только  $|I| = 3037$  из них. Поэтому в алгоритме 2 поиск трикластеров начинается сразу со строки 7 алгоритма 1.