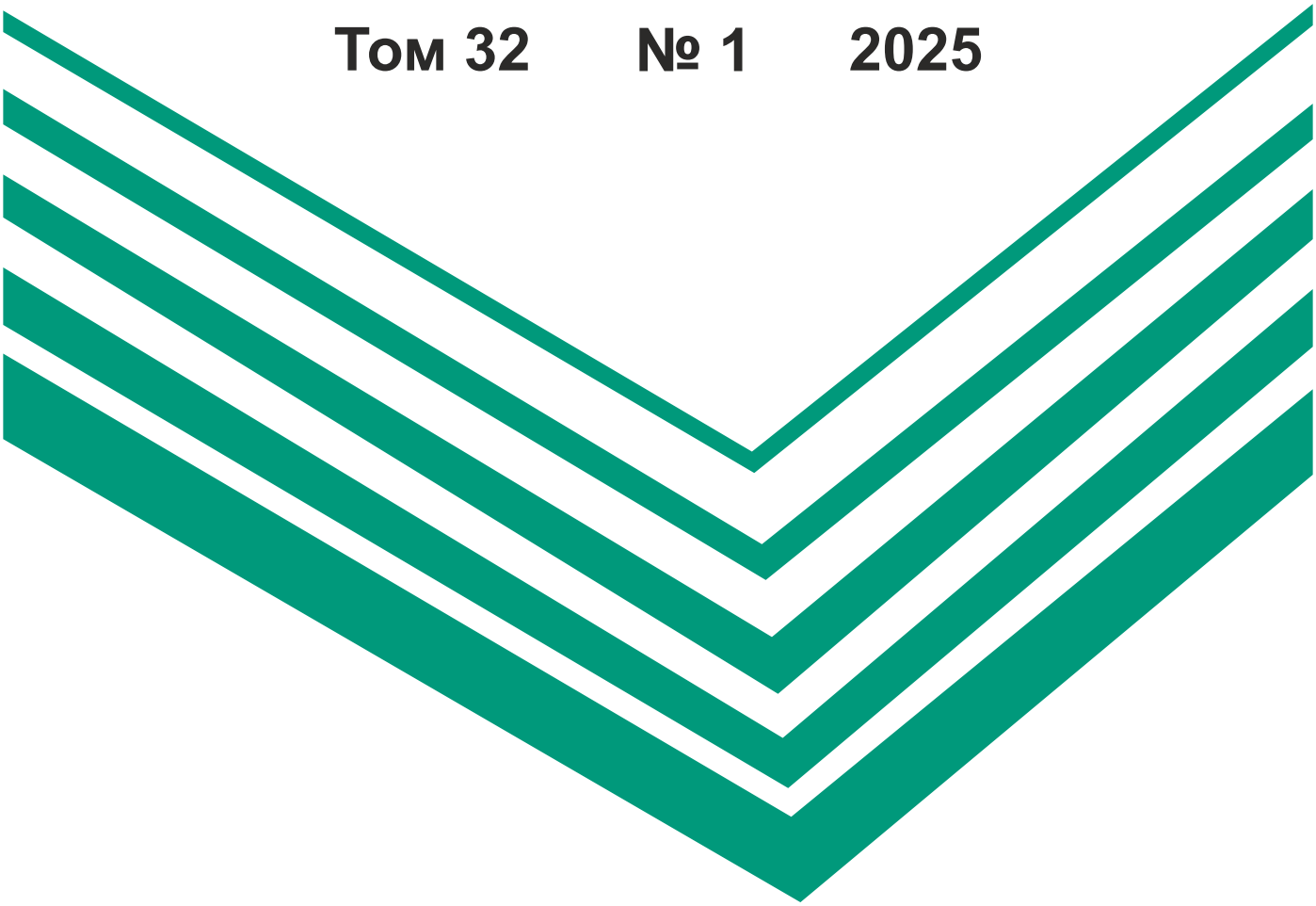
ISSN 2949-5598

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

Том 32 № 1 2025



Новосибирск
Издательство Института математики

ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Январь–март 2025. Т. 32, № 1. С. 5–15

УДК 519.254 DOI: 10.33048/daio.2025.32.812

ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ ЗАДАЧИ ВЫБОРА ТИПИЧНЫХ ПРЕДСТАВИТЕЛЕЙ В КОНЕЧНОМ МНОЖЕСТВЕ ТОЧЕК МЕТРИЧЕСКОГО ПРОСТРАНСТВА

И. А. Борисова[1,2]

[1] Институт математики им. С. Л. Соболева,
пр. Акад. Коптюга, 4, 630090 Новосибирск, Россия
[2] Новосибирский гос. университет,
ул. Пирогова, 2, 630090 Новосибирск, Россия
E-mail: biamia@mail.ru

Аннотация. Исследуется сложностной статус одной экстремальной задачи выбора подмножества p наиболее типичных точек в заданном конечном множества точек метрического пространства. При этом требуется, чтобы выбранное подмножество наилучшим образом описывало исходное множество с точки зрения некоторого геометрического критерия. Рассматриваемая задача является формализацией одной прикладной проблемы из анализа данных, заключающейся в отыскании подмножества типичных представителей выборки с опорой на функцию конкурентного сходства. В статье доказывается, что рассматриваемая задача NP-трудна. Для этого к ней полиномиально сводится одна из NP-трудных задач — задача о трёхмерном сочетании. Ил. 1, библиогр. 14.

Ключевые слова: NP-трудная задача, выбор типичных представителей, функция конкурентного сходства, задача о трёхмерном сочетании, машинное обучение.

Введение

Несмотря на то, что современное общество всё чаще требует от науки сосредоточить усилия на решении актуальных прикладных задач, подобная прагматизация не должна ставить крест на фундаментальных исследованиях, а потому одним из важных вызовов для современной прикладной математики становится формализация и детальное изучение задач, порождаемых практикой.

В данной работе предлагается формальная постановка и устанавливается сложностной статус одной такой задачи, решаемой в процессе

© И. А. Борисова, 2025

интеллектуального анализа данных, задачи выбора типичных объектов из неклассифицированной выборки, максимально сохраняющих информацию о ней, с точки зрения некоторого формального критерия. Содержательно эта задача близка к задаче кластеризации, в которой множество объектов выборки разбивается на кластеры, состоящие из объектов, одновременно максимально похожих на центральные элементы этих кластеров и непохожих на центральные элементы соседних кластеров.

1. Формулировка задачи, её интерпретация и истоки

Задача выбора типичных объектов из неклассифицированной выборки в метрическом пространстве возникает в распознавании образов и машинном обучении при анализе выборок, описание которых необходимо «сжать», переходя от всего множества объектов к кластерам объектов, представленных наиболее типичными их представителями. Так, при анализе профилей клиентской активности сотовые операторы формируют кластеры клиентов со схожими потребностями и создают новые тарифы для типичных клиентов из каждого кластера. Принципы, в соответствии с которыми осуществляется отбор типичных представителей при машинном обучении, могут меняться в зависимости от того, как формализуется понятие типичного объекта.

В рассматриваемой постановке типичными считаются те объекты выборки, на которые максимально похожи объекты из того же кластера и непохожи объекты других кластеров. В качестве меры похожести-непохожести для объектов, представленных в виде точек многомерного метрического пространства, используется функция конкурентного сходства, или FRiS-функция (function of rival similarity) [1], которая позволяет вычислять величину конкурентного сходства на основе расстояний между объектами.

Для множества $\{f_i\}_{i=1}^n$ конечной мощности n через $\min\limits_{i\in\overline{1,n}}{}' f_i$ обозначим его второй минимальный элемент, т. е. если $f_{i_1} \leqslant f_{i_2} \leqslant \ldots \leqslant f_{i_n}$, то $\min\limits_{i\in\overline{1,n}}{}' f_i = f_{i_2}$ для $n \geqslant 2$ и $\min\limits_{i\in\overline{1,n}}{}' f_i = f_1$ для $n = 1$. В этих обозначениях имеем следующую постановку.

Задача 1. *Даны конечное множество точек Λ в метрическом пространстве размерности d и натуральное p. Требуется найти непустое подмножество $S_p \subseteq \Lambda$, которое максимизирует целевую функцию*

$$F(S_p) = \sum_{z\in\Lambda} \frac{\min\limits_{s_2\in S_p}{}' r(z,s_2) - \min\limits_{s_1\in S_p} r(z,s_1)}{\min\limits_{s_2\in S_p}{}' r(z,s_2) + \min\limits_{s_1\in S_p} r(z,s_1)}, \tag{1}$$

где $r(a,b)$ — расстояние между точками a и b, при ограничении $|S_p| = p$.

Исследования сложностного статуса различных задач машинного обучения, использующих в процессе анализа данных функцию конкурентного сходства, ведутся уже несколько лет [2–4].

Наиболее близкая постановка данной задачи встречается в работе [4], в которой принцип выбора подмножества типичных объектов также основан на использовании функции конкурентного сходства, однако в ней автор рассматривает редуцированный критерий, из которого исключён компонент, оценивающий непохожесть объектов на типичных представителей чужих кластеров.

Задача выбора подмножества типичных объектов из конечного множества близка к таким известным задачам кластеризации, как k-means (или k-MSSC) [5, 6] и k-median [7] в их дискретной постановке [8, 9], когда центры кластеров отыскиваются не среди всех точек метрического пространства, а только среди тех, которые включены в исходное множество точек. Похожие постановки есть в [10, 11], где в процессе кластеризации отыскиваются наиболее типичные элементы, в то время как элементы-выбросы в формируемые кластеры не включаются. Несколько вариантов постановки задачи выбора подмножества точек из заданного множества для различных геометрических критериев представлены в [12]. В [2, 13] также рассматривается задача поиска типичных представителей. В этой задаче, в отличие от всех вышеперечисленных, которые можно интерпретировать как поиск центров неизвестных кластеров, разбиение множества точек на два кластера задаётся изначально, причём каждый кластер может описываться несколькими точками, вместо одного центра.

2. Описание математической модели

Теперь поподробнее рассмотрим задачу анализа данных, моделируемую в настоящей статье. Предположим, что для каждой тройки объектов $x, y, z \in A$ задан способ вычислять конкурентное сходство $f(z, x, y)$ таким образом, что чем больше z похож на объект x и непохож на объект y, тем величина $f(z, x, y)$ больше. Тогда типичными с точки зрения конкурентного сходства считаются объекты, у которых сходство с объектами своего кластера как можно выше, а с объектами чужого кластера — как можно ниже.

Если разбиение на кластеры $\{A_i\}_{i=1}^p$ задано изначально, так что

$$A = \bigcup_{i=1}^{p} A_i, \quad \text{где } A_i \cap A_j = \varnothing \text{ для } i \neq j,$$

то задача выбора подмножества p типичных объектов из этой выборки заключается в отыскании такого набора элементов $B = \{b_i\}_{i=1}^p$, $b_i \in A_i$,

которые обеспечат максимальное сходство всех объектов выборки с типичными объектами своего кластера при минимальном сходстве с типичными объектами другого кластера:

$$F_{\text{c}}(B) = \sum_{i=1}^{p} \sum_{z \in A_i} \min_{\substack{y \in B, \\ y \neq b_i}} f(z, b_i, y) \to \max_{\substack{B = \{b_i\}_{i=1}^{p}, \\ b_i \in A_i}}.$$

Если же разбиение на кластеры не задано, то из требования максимального сходства объектов с типичным представителем своего кластера следует, что своим считается тот кластер, сходство с типичным объектом которого максимально. Тогда задача записывается следующим образом:

$$F_{nc}(B) = \sum_{z \in A} \max_{b \in B} \min_{\substack{y \in B, \\ y \neq b}} f(z, b, y) \to \max_{\substack{B \subset A, \\ |B| = p}}. \tag{2}$$

В большинстве случаев при решении прикладных задач каждый объект из множества A рассматривается как точка в пространстве описывающих характеристик, на котором определена некоторая метрика r. Это позволяет вычислять расстояние $r(x, y)$ между парой объектов $x, y \in A$. В рассматриваемой постановке в качестве величины конкурентного сходства $f(z, x, y)$ используется FRiS-функция [1], определяемая по следующей формуле:

$$\text{FRiS}(z, x, y) = \frac{r(z, y) - r(z, x)}{r(z, y) + r(z, x)}.$$

Если подставить формулу для вычисления FRiS-функции в максимизируемый функционал задачи (2), то его слагаемые преобразуются следующим образом:

$$F(B) = \sum_{z \in A_i} \max_{b \in B} \min_{\substack{y \in B, \\ y \neq b}} \frac{r(z, y) - r(z, b)}{r(z, y) + r(z, b)} = \sum_{i=1}^{p} \sum_{z \in A_i} \frac{\min\limits_{\substack{y \in B, \\ y \neq y_z}} r(z, y) - \min\limits_{b \in B} r(z, b)}{\min\limits_{\substack{y \in B, \\ y \neq y_z}} r(z, y) + \min\limits_{b \in B} r(z, b)},$$

где $y_z = \arg\min\limits_{b \in B} r(z, b)$. Для получения критерия (1) из задачи 1 осталось заметить, что

$$\min_{\substack{y \in B, \\ y \neq y_z}} r(z, y) = \min_{y \in B}{}' r(z, y).$$

3. Вычислительная сложность задачи выбора подмножества типичных объектов

Для установления сложностного статуса задачи 1 покажем, что к ней сводится задача о трёхмерном сочетании (3D-matching) в следующей постановке.

Задача 3DM (о трёхмерном сочетании)**.** *Даны три конечных множества* $X = \{x_i\}_{i=1}^{p}$, $Y = \{y_i\}_{i=1}^{p}$, $Z = \{z_i\}_{i=1}^{p}$, *а также множество троек* $T = \{t_i\}_{i=1}^{M} \subset X \times Y \times Z$. *Требуется ответить, существует ли такое подмножество троек* $S \subseteq T$, $|S| = p$, *которое удовлетворяет следующему условию: для любой пары различных троек* $s_1 = (x_1, y_1, z_1)$, $s_2 = (x_2, y_2, z_2) \in S$ *выполняются неравенства* $x_1 \neq x_2$, $y_1 \neq y_2$ *и* $z_1 \neq z_2$.

Задача о трёхмерном сочетании является одной из NP-полных задач из списка Карпа [14].

Теорема 1. *Задача о трёхмерном сочетании полиномиально сводится к задаче* 1.

Доказательство. Рассмотрим задачу 1 для случая, когда исходное множество точек Λ в метрическом пространстве состоит из следующих групп: множеств $\widetilde{X} = \{\widetilde{x}_i\}_{i=1}^{p}$, $\widetilde{Y} = \{\widetilde{y}_i\}_{i=1}^{p}$ и $\widetilde{Z} = \{\widetilde{z}_i\}_{i=1}^{p}$, множества $\widetilde{T} = \{\widetilde{t}_i\}_{i=1}^{M}$ и набора множеств $B_1, B_2, \ldots, B_M$ таких, что $|B_1| = |B_2| = \cdots = |B_M| = N > M + 3$. При этом расстояния между точками задаются следующим образом:

- $r(\widetilde{t}_i, \widetilde{x}_j) = r(\widetilde{t}_i, \widetilde{y}_k) = r(\widetilde{t}_i, \widetilde{z}_l) = 0{,}5$, если $\widetilde{t}_i = (\widetilde{x}_j, \widetilde{y}_k, \widetilde{z}_l) \in \widetilde{T}$;
- $r(\widetilde{t}_i, b) = 0{,}5$ для $\widetilde{t}_i \in \widetilde{T}$ и всех $b \in B_i$;
- $r(x, y) = 1$ для любой другой пары точек $x, y \in \Lambda$.

В этом случае матрица попарных расстояний удовлетворяет аксиомам метрического пространства, в том числе и аксиоме треугольника.

На рис. 1 проиллюстрирован внешний вид полученного множества точек Λ. Ребро между двумя точками означает, что расстояние между ними равно 0,5.

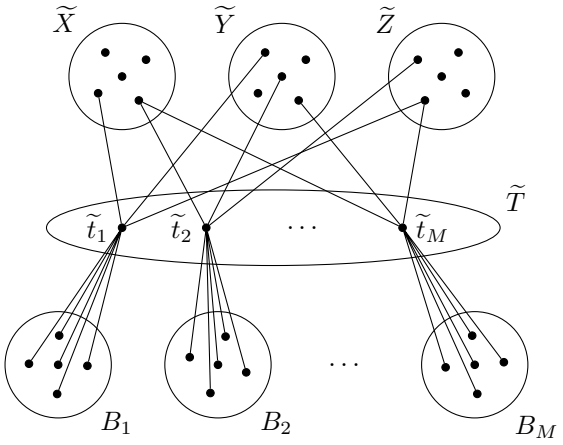


Рис. 1. Структура множества Λ из теоремы 1

Покажем, что только $S \subseteq \widetilde{T}$ может быть решением задачи 1. Докажем это для тривиального случая, когда $p = 1$. При этом по определению $\min\limits'_{s_2 \in S} r(z, s_2) = 1$. Если $S_1 = \{b\}$, где $b \in B_i$, то расстояние от b только до одного t_i равно 0,5, поэтому

$$F(\{b\}) = \sum_{z \in \Lambda} \frac{1 - r(z, b)}{1 + r(z, b)} = \frac{1}{3} + 1 = 1\frac{1}{3}.$$

Если $S_1 = \{x\}$, где $x \in \widetilde{X}$, то только расстояния от x до l троек из множества $\widetilde{T}$, содержащих x, равны 0,5, поэтому

$$F(\{x\}) = \sum_{z \in \Lambda} \frac{1 - r(z, x)}{1 + r(z, x)} = l \cdot \frac{1}{3} + 1 \leqslant M \cdot \frac{1}{3} + 1.$$

Если $S_1 = \{y\}$, где $y \in \widetilde{Y}$, то по аналогии с предыдущим случаем

$$F(\{y\}) \leqslant M \cdot \frac{1}{3} + 1.$$

Если $S_1 = \{z\}$, где $z \in \widetilde{Z}$, то по аналогии с предыдущим случаем

$$F(\{z\}) \leqslant M \cdot \frac{1}{3} + 1.$$

Если $S_1 = \{t\}$, где $t = (x, y, z) \in \widetilde{T}$, то только расстояния от t до N точек связанного с ним множества B_i и до x, y, z равны 0,5, поэтому

$$F(\{t\}) = \sum_{z \in \Lambda} \frac{1 - r(z, t)}{1 + r(z, t)} = (N + 3) \cdot \frac{1}{3} + 1 > (M + 3 + 3) \cdot \frac{1}{3} + 1 = M \cdot \frac{1}{3} + 3.$$

Таким образом, максимум функционала достигается на множестве столпов S_1, состоящем из элемента $t \in \widetilde{T}$.

Теперь докажем, что при дальнейшем увеличении числа элементов в S по-прежнему максимальное значение функционала качества будет достигаться на элементах из $\widetilde{T}$.

Пусть уже имеется множество столпов $S_{k-1} \subset \widetilde{T}$, являющееся решением задачи 1 для $p = k - 1$. Оценим, как меняется качество при добавлении в S_{k-1} элементов из разных подмножеств Λ. Для $z \in \Lambda$ введём обозначение

$$F(z, S) = \frac{\min\limits'_{s_2 \in S} r(z, s_2) - \min\limits_{s_1 \in S} r(z, s_1)}{\min\limits'_{s_2 \in S} r(z, s_2) + \min\limits_{s_1 \in S} r(z, s_1)}.$$

Если $S_k = S_{k-1} \cup \{b\}$, где $b \in B_i$, то расстояние от b до t_i равно 0,5 и добавление b в S_k повлияет только на значения $F(b, S_k)$ и $F(t_i, S_k)$, поэтому $F(S_k) \leqslant F(S_{k-1}) + 1 + \frac{1}{3}$.

Если $S_k = S_{k-1} \cup \{x\}$, где $x \in \widetilde{X}$, то расстояния от x до l троек из множества $\widetilde{T}$, содержащих x, равны 0,5 и добавление x повлияет на значение $F(x, S_k)$ и l значений вида $F(t_i, S_k)$, поэтому

$$F(S_k) \leqslant F(S_{k-1}) + 1 + l \cdot \frac{1}{3} \leqslant F(S_{k-1}) + 1 + M \cdot \frac{1}{3}.$$

Аналогичные неравенства выполняются для $y \in \widetilde{Y}$ и $z \in \widetilde{Z}$.

Если $S_k = S_{k-1} \cup \{t\}$, где $t = (x, y, z) \in \widetilde{T}$, то расстояния от t до N элементов связанного с ним множества B_i и до x, y, z равны 0,5, и добавление t повлияет на значение $F(x, S_k)$ и N значений вида $F(b, S_k)$ для $b \in B_i$. При этом для $b \in B_i$ оно увеличится с 0 до $\frac{1}{3}$, для t — с 0 до 1. Значения $F(x, S_k)$, $F(y, S_k)$ $F(z, S_k)$ могут либо увеличиться с 0 до $\frac{1}{3}$, либо остаться равными 0, либо уменьшиться с $\frac{1}{3}$ до 0 в зависимости от того, были ли в наборе S_{k-1} тройки, содержащие x, y, z соответственно. В результате имеем

$$F(S_k) \geqslant F(S_{k-1}) + 1 + N \cdot \frac{1}{3} - 3 \cdot \frac{1}{3} > F(S_{k-1}) + M \cdot \frac{1}{3} + 1.$$

Таким образом, доказано, что множество столпов S_k может быть образовано только элементами $t \in \widetilde{T}$ при любом $k \leqslant M$.

Осталось доказать, что множество S_p, являющееся решением задачи 1, за линейное время может быть преобразовано в решение задачи 3DM S путём замены каждого $\widetilde{t}_i \in \widetilde{T}$ на $t_i \in T$ с тем же индексом.

Предположим, что множество S, полученное таким образом, не является решением задачи 3DM, т. е. в нём существует пара различных троек, пересекающихся минимум по одному элементу. При этом существует множество S^*, которое является решением задачи 3DM и обеспечивает условие непересекаемости всех троек. Рассмотрим образ этого множества S_p^*, полученный переходом от множества объектов из T к множеству объектов из $\widetilde{T}$ с теми же индексами.

Сравним качество $F(S_p)$ и $F(S_p^*)$. Так как каждый из p элементов множеств $\widetilde{X}$, $\widetilde{Y}$ и $\widetilde{Z}$ имеет ровно один элемент из S_p^* на расстоянии 0,5, значение FRiS-функции для них на этом множестве типичных представителей равно $\frac{1}{3}$. Аналогично для N каждого из p множеств B_i, связанных с теми t_i, которые попали в S_p^*. Для самих $t_i \in S_p^*$ значения FRiS-функции равны 1. В результате имеем

$$F(S_p^*) = 3 \cdot p \cdot \frac{1}{3} + N \cdot p \cdot \frac{1}{3} + p = 2p + \frac{1}{3}Np.$$

С другой стороны, если в множестве столпов присутствует пара образов пересекающихся троек, то величина FRiS-функции для элемента из $\widetilde{X}$, $\widetilde{Y}$ или $\widetilde{Z}$, по которому они пересекаются, равна 0, так как два ближайших

к нему столпа находятся на одинаковом расстоянии 0,5. Тогда

$$F(S_p) \leqslant (3p-1)\cdot\frac{1}{3} + Np\cdot\frac{1}{3} + p < F(S_p^*),$$

а следовательно, S_p решением задачи 1 быть не может.

Таким образом, если $F(S_p) = 2p+\frac{1}{3}Np$, то задача 3DM имеет решение, и получить его можно, включив в S образы всех элементов из S_p. Если же $F(S_p) < 2p + \frac{1}{3}Np$, то решения задачи 3DM не существует. Теорема 1 доказана.

Следствие 1. *Задача* 1 *выбора подмножества из множества точек в метрическом пространстве* NP*-трудна.*

Доказательство. Следствие справедливо в силу того, что к рассматриваемой задаче за полиномиальное время сводится задача о трёхмерном сочетании, которая NP-трудна. Следствие 1 доказано.

Заключение

В работе доказано, что известная NP-трудная в сильном смысле задача о трёхмерном сочетании сводится к задаче выбора из конечного множества точек метрического пространства подмножества p типичных представителей, наилучшим образом описывающих кластеры этого множества с точки зрения некоторого геометрического критерия. Тем самым доказано, что указанная задача выбора NP-трудна, и в рамках парадигмы $\mathrm{P} \neq \mathrm{NP}$ невозможно предложить точный полиномиальный алгоритм решения этой задачи.

Финансирование работы

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект № 17–01–00710). Дополнительных грантов на проведение или руководство этим исследованием получено не было.

Конфликт интересов

Автор заявляет, что у неё нет конфликта интересов.

Литература

1. **Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A.** Methods of recognition based on the function of rival similarity // Pattern Recognit. Image Anal. 2008. V. 18, No. 1. P. 1–6.
2. **Борисова И. А.** Вычислительная сложность задачи выбора типичных представителей в 2-разбиении конечного множества точек метрического пространства // Дискрет. анализ и исслед. операций. 2020. Т. 27, № 2. С. 5–16.

3. **Кутненко О. А., Плясунов А. В.** NP-трудность некоторой задачи цензурирования данных // Дискрет. анализ и исслед. операций. 2021. Т. 28, № 2. С. 60–73.
4. **Кутненко О. А.** Вычислительная сложность двух задач когнитивного анализа данных // Дискрет. анализ и исслед. операций. 2022. Т. 29, № 1. С. 18–32.
5. **Aloise D., Deshpande A., Hansen P., Popat P.** NP-hardness of Euclidean sum-of-squares clustering // Machine Learn. 2009. V. 75, No. 2. P. 245–248.
6. **Dasgupta S.** The hardness of k-means clustering. Tech. Rep. CS2007-0890. San Diego, CA: Univ. California, 2008. 6 p.
7. **Papadimitriou C. H.** Worst-case and probabilistic analysis of a geometric location problem // SIAM J. Comput. 1981. V. 10, No. 3. P. 542–557.
8. **Har-Peled S., Mazumdar S.** On coresets for k-means and k-median clustering and their applications // Proc. 36th Annu. ACM Symp. Theory of Computing (Chicago, USA, June 13–16, 2004). New York: ACM, 2004. P. 291–300.
9. **Kaufman L., Rousseeuw P. J.** Clustering by means of medoids // Statistical data analysis based on the L_1-norm and related methods. Amsterdam: North Holland, 1987. P. 405–416.
10. **Кельманов А. В., Пяткин А. В., Хандеев В. И.** О сложности некоторых максиминных задач кластеризации // Тр. Ин-та математики и механики. 2018. Т. 24, № 4. С. 189–198.
11. **Кельманов А. В., Пяткин А. В.** NP-трудность некоторых евклидовых задач разбиения конечного множества точек // Журн. вычисл. математики и мат. физики. 2018. Т. 58, № 5. С. 852–856.
12. **Aggarwal H., Imai N., Katoh N., Suri S.** Finding k points with minimum diameter and related problems // J. Algorithms. 1991. V. 12, No. 1. P. 38–56.
13. **Zukhba A. V.** NP-completeness of the problem of prototype selection in the nearest neighbor method // Pattern Recognit. Image Anal. 2010. V. 20, No. 4. P. 484–494.
14. **Karp R. M.** Reducibility among combinatorial problems // Complexity of computer computations. Proc. Symp. (New York, USA, March 20–22, 1972). New York: Plenum Press, 1972. P. 85–103.

Борисова Ирина Артёмовна

Статья поступила
3 сентября 2024 г.
После доработки —
16 сентября 2024 г.
Принята к публикации
22 сентября 2024 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
January–March 2025. Vol. 32, No. 1. P. 5–15

UDC 519.254 DOI: 10.33048/daio.2025.32.812

COMPUTATIONAL COMPLEXITY OF THE CHOICE PROBLEM FOR TYPICAL REPRESENTATIVES OF A FINITE POINT SET IN A METRIC SPACE

I. A. Borisova[1,2]

[1] Sobolev Institute of Mathematics,
4 Acad. Koptyug Avenue, 630090 Novosibirsk, Russia
[2] Novosibirsk State University,
2 Pirogov Street, 630090 Novosibirsk, Russia
E-mail: biamia@mail.ru

Abstract. We analyze the complexity of one extremal problem of choosing a subset of p points in a given finite set in a metric space. The chosen subset of points is required to describe given clusters in the best way from the point of view of some geometric criterion. This problem is a formalization of one applied problem from data mining that consists in finding a subset of typical representatives of a dataset based on the rival similarity function of. We prove that the problem under consideration is NP-hard by polynomially reducing the well-known NP-hard 3D-Matching problem to this one. Illustr. 1, bibliogr. 14.

Keywords: NP-hard problem, choice of typical representatives, FRiS-function, 3D-matching problem, machine learning.

References

1. **N. G. Zagoruiko, I. A. Borisova, V. V. Dyubanov,** and **O. A. Kutnenko,** Methods of recognition based on the function of rival similarity, *Pattern Recognit. Image Anal.* **18** (1), 1–6 (2008).
2. **I. A. Borisova,** Computational complexity of the problem of choosing typical representatives in a 2-clustering of a finite set of points in a metric space, *Diskretn. Anal. Issled. Oper.* **27** (2), 5–16 (2020) [Russian] [*J. Appl. Ind. Math.* **14** (2), 242–248 (2020)].
3. **O. A. Kutnenko** and **A. V. Plyasunov,** NP-hardness of some data cleaning problem, *Diskretn. Anal. Issled. Oper.* **28** (2), 60–73 (2021) [Russian] [*J. Appl. Ind. Math.* **15** (2), 285–291 (2021)].

English transl.: *Journal of Applied and Industrial Mathematics* **19** (1), 1–6 (2025), DOI: 10.1134/S1990478925010016.

4. **O. A. Kutnenko,** Computational complexity of two problems of cognitive data analysis, *Diskretn. Anal. Issled. Oper.* **29** (1), 18–32 (2022) [Russian] [*J. Appl. Ind. Math.* **16** (1), 89–97 (2022)].
5. **D. Aloise, A. Deshpande, P. Hansen,** and **P. Popat,** NP-hardness of Euclidean sum-of-squares clustering, *Machine Learn.* **75** (2), 245–248 (2009).
6. **S. Dasgupta,** The hardness of k-means clustering, *Tech. Rep. CS2007-0890* (Univ. California, San Diego, CA, 2008).
7. **C. H. Papadimitriou,** Worst-case and probabilistic analysis of a geometric location problem, *SIAM J. Comput.* **10** (3), 542–557 (1981).
8. **S. Har-Peled** and **S. Mazumdar,** On coresets for k-means and k-median clustering and their applications, in *Proc. 36th Annu. ACM Symp. Theory of Computing, Chicago, USA, June 13–16, 2004* (ACM, New York, 2004), pp. 291–300.
9. **L. Kaufman** and **P. J. Rousseeuw,** Clustering by means of medoids, in *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (North Holland, Amsterdam, 1987), pp. 405–416.
10. **A. V. Kel'manov, A. V. Pyatkin,** and **V. I. Khandeev,** On the complexity of some max–min clustering problems, *Tr. Inst. Mat. Mekh.* **24** (4), 189–198 (2018) [Russian] [*Proc. Steklov Inst. Math.* **309** (Suppl. 1), S65–S73 (2020)].
11. **A. V. Kel'manov** and **A. V. Pyatkin,** NP-hardness of some Euclidean problems of partitioning a finite set of points, *Zh. Vychisl. Mat. Mat. Fiz.* **58** (5), 852–856 (2018) [Russian] [*Comput. Math. Math. Phys.* **58** (5), 822–826 (2018)].
12. **H. Aggarwal, N. Imai, N. Katoh,** and **S. Suri,** Finding k points with minimum diameter and related problems, *J. Algorithms* **12** (1), 38–56 (1991).
13. **A. V. Zukhba,** NP-completeness of the problem of prototype selection in the nearest neighbor method, *Pattern Recognit. Image Anal.* **20** (4), 484–494 (2010).
14. **R. M. Karp,** Reducibility among combinatorial problems, in *Complexity of Computer Computations* (Proc. Symp., New York, USA, March 20–22, 1972) (Plenum Press, New York, 1972), pp. 85–103.

Irina A. Borisova

Received September 3, 2024
Revised September 16, 2024
Accepted September 22, 2024