

Ю. А. Мезенцев, д-р техн. наук, проф., e-mail: mesyan@yandex.ru,
О. М. Разумникова, д-р биол. наук, проф., e-mail: razoum@mail.ru,
Новосибирский государственный технический университет,
И. В. Тарасова, д-р мед. наук, вед. науч. сотр., e-mail: iriz78@mail.ru,
О. А. Трубникова, д-р мед. наук, зав. лабораторией, e-mail: olgalet17@mail.ru,
ФГБНУ НИИ комплексных проблем сердечно-сосудистых заболеваний, г. Кемерово

О некоторых задачах кластеризации больших данных по минимаксным и аддитивным критериям, применение в медицине и нейрофизиологии*

Рассматриваются NP-трудные задачи кластеризации в применении к данным нейрофизиологических исследований (показателей послеоперационной когнитивной дисфункции). Приводятся варианты постановок задач кластеризации в виде задач смешанного целочисленного программирования, в том числе с использованием непрерывной релаксации, снижающей трудоемкость решений без потери точности. Представлены результаты вычислительных экспериментов с серией реализаций поставленных задач на реальных данных, с использованием программной реализации алгоритма бинарных отсечений и ветвлений, демонстрирующие высокую эффективность разработанного инструментария.

Ключевые слова: кластеризация, минимаксный критерий качества, аддитивный критерий, линейная релаксация, алгоритм бинарных отсечений и ветвлений, выявление послеоперационной когнитивной дисфункции

Введение

Различные модификации прикладных задач кластеризации и соответствующих алгоритмов решения известны довольно давно. Этой проблематике посвящено множество работ по прикладной статистике [1], методам дискретной оптимизации [2, 3]. Большинство подобных постановок относятся к разряду труднорешаемых задач смешанного целочисленного программирования [2]. Настоящая работа посвящена решению ряда модификаций прикладной задачи кластеризации, выявлению и сравнению содержательных закономерностей разбиений различных множеств объектов на подмножества с сопоставлением таких разбиений по частотным характеристикам электрической активности мозга у пациентов до коронарного шунтирования и через 7...10 суток после операции.

*Работа поддержана Российским фондом фундаментальных исследований, проект № 19-29-01017.

1. Постановка задачи кластеризации по критерию минимума максимальной суммы расстояний между ближайшими парами объектов внутри кластера

В данной статье развивается подход, основанный на применении инструментария (моделей и методов) смешанного целочисленного линейного программирования (milp). Кластеризация осуществляется посредством решения специальных задач названного класса, которые, в терминах milp, можно интерпретировать как разновидности задачи о назначениях исходного множества объектов на множество кластеров. При интерпретации в терминах теории расписаний объекты рассматриваются в качестве заявок, а кластеры — в качестве приборов, на которые назначаются заявки. Подобная задача сформулирована в работах [4, 5]. Для каждой такой заявки и каждого прибора определены характеристики переходов от любых текущих к последующим заявкам (цена, время, коэффициент корреляции, эвклидово

расстояние, либо любой другой скаляр, который для простоты будем именовать расстоянием). Вполне очевидна связь данной задачи с задачами маршрутизации (vgr). В этом контексте последовательности переходов между объектами не фиксированы и подлежат определению. Пересечения кластеров по объектам (и объектам по кластерам) не допускаются.

Для формального представления задачи введем следующие обозначения:

$i, j = \overline{1, n}$ — номера объектов; $l, k = \overline{1, m}$ — номера кластеров; $c_{i,j}^k$ — расстояния между объектами i и j в кластере k . Тогда задача кластеризации состоит в определении булевых переменных $x_{i,j}^k$ при выполнении ряда условий:

$$x_{i,j}^k = \begin{cases} 1, & \text{если в кластер } k \text{ включается} \\ & \text{пара объектов } i \text{ и } j; \\ 0 & \text{в противном случае,} \end{cases} \quad (1)$$

$$\sum_{k=1}^m \sum_{i=1}^n x_{i,j}^k = 1, \quad j = \overline{1, n}, \quad i \neq j; \quad (2)$$

$$\sum_{k=1}^m \sum_{j=1}^n x_{i,j}^k = 1, \quad i = \overline{1, n}, \quad i \neq j; \quad (3)$$

$$\sum_{k=1, k \neq l}^m \sum_{j=1}^n x_{i,j}^k + \sum_{j=1}^n x_{j,i}^l \leq 1, \quad i = \overline{1, n}, \quad l = \overline{1, m}; \quad (4)$$

$$\sum_{i=1}^n x_{j,i}^l + \sum_{k=1, k \neq l}^m \sum_{i=1}^n x_{i,j}^k \leq 1, \quad j = \overline{1, n}, \quad l = \overline{1, m}; \quad (5)$$

$$\sum_{j=1}^n \sum_{i=1}^n r_{j,s}^k x_{i,j}^k \leq R_s^k, \quad k = \overline{1, m}, \quad s = \overline{1, S}; \quad (6)$$

$$\sum_{j=1}^n \sum_{i=1}^n c_{i,j}^k x_{i,j}^k \leq \lambda, \quad k = \overline{1, m}; \quad (7)$$

$$\lambda \rightarrow \min. \quad (8)$$

Ограничения-равенства (2)—(3) с учетом (1) являются обычными условиями задачи о назначениях (любой объект может принадлежать только одному кластеру).

В совокупности (4)—(5) означают следующее. Сумма числа назначений переходов от объектов-источников для всех кластеров, кроме кластера l , к объектам-стокам кластера l меньше либо равна единице. Данные условия запрещают принадлежность объектов всем кластерам, кроме кластера l . Сумма числа назначений переходов от объектов-источников для кластера l к объектам-стокам всех кластеров, кроме кластера l , меньше либо равна единице. Это запрещает передачу объектов, назначенных кластеру l , любым другим кластерам. В этом

контексте предполагается определение последовательностей переходов от одних объектов внутри кластера к другим. Этим же обусловлено использование понятий объекта-источника и объекта-стока, которые в совокупности определяют отношения предшествования—следования между объектами, которые, в частности, используются в задачах vgr.

Задачи кластеризации могут дополняться ресурсными ограничениями вида (6). Чаще всего посредством ресурсных ограничений учитываются пропускные способности коммуникаций и грузоподъемности транспортных средств в задачах маршрутизации.

Условия (7) и (8) в совокупности реализуют минимаксный критерий, интерпретируемый как критерий быстродействия рассматриваемой параллельно-последовательной обслуживаемой системы и как критерий равномерной нагрузки на кластеры [6, 7].

2. Постановка задачи кластеризации по критерию минимума максимальной суммы расстояний между всеми парами объектов внутри кластера

Задачи кластеризации, не связанные с маршрутизацией, требуют несколько иного подхода к определению эффективности принадлежности (назначений) объектов кластерам.

Наиболее рациональной и чаще всего используемой мерой является сумма расстояний между всеми парами объектов внутри кластера [1]. Ее использование существенно усложняет формальную постановку задачи в сравнении с (1)—(8). Приведем формальную постановку задачи оптимальной кластеризации для этого случая.

Определим переменные y_i^k (идентифицирующие принадлежность объектов $i, j, i, j = \overline{1, n}$, кластеру $k, k = \overline{1, m}$) при выполнении ряда условий:

$$y_i^k = \begin{cases} 1, & \text{если объект } i \text{ принадлежит} \\ & \text{кластеру } k; \\ 0 & \text{в противном случае, } i = \overline{1, n}; \end{cases} \quad (9)$$

$$\sum_{k=1}^m y_i^k = 1, \quad i = \overline{1, n}. \quad (10)$$

Определим также зависимые переменные $x_{i,j}^k = y_i^k y_j^k$ и линеаризующие неравенства [8]:

$$0 \leq y_i^k + y_j^k - 2x_{i,j}^k \leq 1, \quad k = \overline{1, m}, \quad i, j = \overline{1, n}, \quad i \neq j,$$

которые при несимметричной матрице расстояний преобразуются в неравенства

$$0 \leq y_i^k + y_j^k - x_{i,j}^k - x_{j,i}^k \leq 1, \quad k = \overline{1, m} \quad i, j = \overline{1, n}, \quad i \neq j, \quad (11)$$

$$x_{i,j}^k = \begin{cases} 1, & \text{если кластеру } k \text{ принадлежат} \\ & \text{объекты } i, j: y_i^k = 1, y_j^k = 1; \\ 0 & \text{в противном случае,} \\ & i, j = \overline{1, n}, \quad i \neq j. \end{cases} \quad (12)$$

Добавив в задачу условия (7) и (8), реализующие минимаксный критерий: $\sum_{j=1}^n \sum_{i=1}^n c_{i,j} x_{i,j}^k \leq \lambda$, $k = \overline{1, m}$, $i \neq j$, $\lambda \rightarrow \min$, на имеющие, однако, несколько иной смысл, чем в задаче (1)–(8) (минимизации максимальной по всем кластерам суммы расстояний между всеми объектами каждого кластера), получим искомую формализацию (7)–(12) названной в заголовке раздела задачи.

Кроме критерия (7)–(8), в зависимости от содержательного смысла задачи кластеризации в ряде случаев более приемлемым является аддитивный критерий, который удобно представлять в виде

$$\sum_{j=1}^n \sum_{i=1}^n c_{i,j} x_{i,j}^k = \lambda^k, \quad k = \overline{1, m}, \quad i \neq j; \quad (13)$$

$$\sum_{k=1}^m \lambda^k \rightarrow \min. \quad (14)$$

Здесь λ^k — сумма расстояний между всеми парами объектов в кластере k , $k = \overline{1, m}$.

Решение задачи (9)–(14) позволяет находить разбиение множества объектов с заданными расстояниями между всеми парами объектов на заданное число (m) подмножеств (кластеров), которое гарантирует минимизацию суммы минимальных суммарных расстояний между всеми парами объектов по всем кластерам.

Оценим трудоемкость вариантов задачи кластеризации (1)–(8), (7)–(12), (9)–(14).

3. Оценки вычислительной сложности задач кластеризации и возможностей релаксации

Формального доказательства NP-трудности представленных постановок задач нами не приводится, поскольку такое доказательство предполагает сведение представленных постановок к любой известной задаче с NP-полнотой, что само по себе может оказаться труднорешаемой

задачей. Заметим только, что при значительном упрощении ограничивающих условий любой из приведенных выше задач получается NP-трудная задача смешанного целочисленного программирования. В частности, подзадачи (1)–(3), (7), (8), либо (7)–(10), (12) интерпретируются как NP-трудная задача оптимизации расписаний несвязанных параллельных машин по критерию C_{\max} . Доказательство ее NP-трудности можно найти, например, в работах [6, 7]. С учетом же остальных условий сложность представленных задач увеличивается на много порядков, что показывают вычислительные эксперименты на реальных данных.

Покажем, как можно несколько ослабить трудоемкость задач (7)–(12) и (9)–(14).

Для этого используем релаксацию по вспомогательным булевым переменным $x_{i,j}^k$, сняв условия целочисленности (12). Вместо них введем границы изменения непрерывных переменных:

$$0 \leq x_{i,j}^k \leq 1, \quad k = \overline{1, m}, \quad i, j = \overline{1, n}, \quad i \neq j. \quad (15)$$

Релаксации задач (7)–(12) и (9)–(14) обозначим соответственно как (7)–(11), (15) и (9)–(11), (13)–(15).

Отмечаем снижение числа булевых переменных в релаксированных задачах на величину mn^2 . Таким образом, общее число булевых переменных (9) в задачах (7)–(11), (15) и (9)–(11), (13)–(15) составляет величину mn при наличии $mn^2 + 1$ непрерывных переменных (15) и (8), против $mn(1 + n)$ булевых переменных в задачах (7)–(12) и (9)–(14). Разница весьма существенна при применении в практических приложениях представленных задач. Поскольку перспективы разработки приемлемых по точности аппроксимационных эффективных алгоритмов для сформулированных задач весьма туманны, успешность практического применения условно экспоненциальных алгоритмов сильно зависит от фактического числа целочисленных переменных. В этом смысле релаксации (7)–(11), (15) и (9)–(11), (13)–(15) имеют существенные преимущества перед постановками (7)–(12) и (9)–(14).

Докажем эквивалентность этих постановок относительно оптимальных значений переменных y_i^k .

Рассмотрим соотношения $x_{i,j}^k = y_i^k y_j^k$ и эквивалентные условия (11) $0 \leq y_i^k + y_j^k - x_{i,j}^k - x_{j,i}^k \leq 1$, $k = \overline{1, m}$, $i, j = \overline{1, n}$, $i \neq j$, означающие, что $x_{i,j}^k$ истинны только тогда, когда истинны y_i^k и y_j^k . Соответственно y_i^k и y_j^k истинны одновременно только тогда, когда истинны $x_{i,j}^k$.

Заменяем условие целочисленности $x_{i,j}^k$ на условие $0 \leq x_{i,j}^k \leq 1$ и рассмотрим возможные варианты соотношений (11): если в оптимальном решении $y_i^k = 1$ и $y_j^k = 1$, то соотношение $x_{i,j}^k = y_i^k y_j^k$ выполнится только в случае $x_{i,j}^k = x_{j,i}^k = 1$. Если в оптимальном решении $y_i^k = 1$ и $y_j^k = 0$, то соотношение $x_{i,j}^k = y_i^k y_j^k$ выполнится только в случае $x_{i,j}^k = x_{j,i}^k = 0$. Совершенно аналогично для случая $y_i^k = 0$ и $y_j^k = 0$ то же соотношение выполнится при $x_{i,j}^k = x_{j,i}^k = 0$. Таким образом, решения релаксаций (7)–(11), (15) и (9)–(11), (13)–(15) совпадают с решениями задач в исходных постановках (7)–(12) и (9)–(14) соответственно.

4. Алгоритмы решения

Для представленных выше вариантов NP-трудных задач кластеризации не существует теоретически эффективных алгоритмов. Однако для практических применений в достаточной степени разработаны алгоритмы, которые можно именовать как условно экспоненциальные. Это словосочетание означает, что, несмотря на недоказанность их эффективности, данные алгоритмы позволяют за разумное время находить оптимальные, либо приближенные к оптимальным решения труднорешаемых задач. Примером такого алгоритма может служить алгоритм бинарных отсечений и ветвлений (АБОВ) [9, 10]. Его программная реализация была применена для поиска решения сформулированных выше задач кластеризации (7)–(11), (15) и (9)–(11), (13)–(15). Ниже приводится содержательная интерпретация результатов обработки данных медико-психологических исследований на основе формальных постановок (7)–(11), (15) и (9)–(11), (13)–(15) и указанного программного приложения.

5. Применение разработанного инструментария кластеризации для выявления послеоперационной когнитивной дисфункции

Серия вычислительных экспериментов кластеризации пациентов на основе представленных выше постановок (7)–(11), (15) и (9)–(11), (13)–(15) и соответствующих алгоритмов дискретной оптимизации [9, 10] выявила высокую эффективность данного инструментария. Задача (7)–(11), (15) с использованием минимаксного критерия и адаптированный под нее алгоритм АБОВ были применены для анализа

показателей нейрофизиологического статуса пациентов, перенесших прямую реваскуляризацию миокарда в условиях искусственного кровообращения. Нейрофизиологическое исследование проводилось за 3...5 дней до операции и через 7...10 суток после коронарного шунтирования. Подробное описание выборки пациентов и регистрации ЭЭГ и показателей когнитивного статуса представлено ранее [11, 12]. Для кластеризации на три группы были использованы суммарные (логарифмированные) значения фоновой мощности ЭЭГ при закрытых глазах в диапазонах $\theta 1$ (4...6 Гц), $\alpha 1$ (8...10 Гц) и $\beta 2$ (20...30 Гц). Эти диапазоны были рассмотрены в связи с их функциональным значением для процессов внимания и памяти [13, 14] и выявленной ранее информационной ролью для предсказания когнитивного снижения в послеоперационном периоде коронарного шунтирования [12]. С учетом большого времени вычислений для анализа была взята выборка из 40 пациентов мужчин ($56,7 \pm 5,08$ лет), имеющих показатели ЭЭГ до и после операции и исходные послеоперационные показатели когнитивного статуса: сумма баллов по краткой шкале оценки психического статуса (КШОПС) и комплексный показатель когнитивного статуса (ККС), являющийся суммарной характеристикой функций произвольного внимания, кратковременной памяти и исполнительных функций [15].

Параметры трех кластеров, полученных в ходе реализации разработанного алгоритма для показателей мощности $\theta 1$, $\alpha 1$ и $\beta 2$ ритма до и после операции, представлены в табл. 1.

Исходя из представленных в табл. 1 результатов можно отметить, что вариабельность состава кластеров составила 33,5 %; 22,5 % и 52,5 %, соответственно для $\theta 1$ -, $\alpha 1$ - и $\beta 2$ -диапазонов. При этом состав пациентов из второго кластера перешел в третий по классификации послеоперационной мощности $\alpha 1$ -биопотенциалов.

Для выяснения функционального значения выделенных кластеров было проведено сравнение значений мощности ЭЭГ в соответствующих частотных диапазонах, возраста и показателей КШОПС и ККС. Однофакторный дисперсионный анализ (ANOVA) был выполнен для каждого из перечисленных показателей с переменной КЛАСТЕР (3)¹. Достоверные эф-

¹Переменная КЛАСТЕР является независимой в однофакторном дисперсионном анализе, показывая, как принадлежность к одному из трех кластеров изменяет показатели мощности $\theta 1$ -, $\alpha 1$ - и $\beta 2$ -ритма.

Таблица 1

Кластеры, сформированные на минимаксном критерии классификации показателей мощности $\theta 1$ -, $\alpha 1$ - и $\beta 2$ -ритма до и после операции коронарного шунтирования

До операции			После операции			
№ кластера	Оценка метрики	<i>n</i>	№ кластера	Оценка метрики	<i>n</i>	<i>N_v</i>
$\theta 1$ -диапазон						
1	45,58	15	1	56,18	17 (10)	(3)2к и (4)3к
2	46,07	14	2	41,04	13 (10)	(3)1к
3	35,51	11	3	43,21	10 (7)	(2)1к и (1)2к
$\alpha 1$ -диапазон						
1	77,05	13	1	80,88	14 (11)	(1)2к и (2)3к
2	83,89	14	2	78,90	14 (4)	(10)3к
3	74,46	13	3	73,86	12 (1)	(1)1к и (10)2к
$\beta 2$ -диапазон						
1	47,80	13	1	48,27	14 (7)	(1)2к и (6)3к
2	48,99	12	2	46,60	12 (7)	(5)3к
3	48,22	15	3	48,12	14 (5)	(6)1к и (1)2к
<p><i>Примечание:</i> <i>n</i> — число пациентов в кластере; в скобках показано число пациентов, оставшихся в этом же кластере после операции или переместившихся из других кластеров, соответственно (<i>N_v</i>).</p>						

факты такого анализа приведены в табл. 2, их post-hoc анализ выполняли с поправкой Бонферрони. Полученные результаты указывают на значимые различия показателей мощности $\theta 1$ -, $\alpha 1$ - и $\beta 2$ -ритма между тремя кластерами как до операции (д/о), так и после нее (п/о). Эффект фактора "Возраст" был обусловлен тем, что кластер 3 характеризовался более молодым возрастом пациентов, а "ККС" — его меньшим значением в группе из кластера 2 (табл. 2).

ANOVA для переменной СТАБИЛЬНОСТЬ (2) (т. е. одну группу составили пациенты из одного и того же кластера, сформированного на основании до- и послеоперационной ЭЭГ; другую — с перемещением между кластерами), выявил достоверный эффект только для ККС и ЭЭГ $\beta 2$ -диапазона: $F(1,38) = 6,170$; $p < 0,018$ с большими значениями ККС в группе с постоянной принадлежностью к кластерам по сравнению с переменной (0,592 и 0,473, соответственно).

Таким образом, можно заключить, что разработанный способ кластеризации перемен-

ных обладает хорошими дискриминационными возможностями, так как сформированные кластеры различаются по заданному критерию с высокой степенью достоверности.

Что касается функционального значения полученных кластеров, то вопрос этот требует дальнейшего исследования. Однако выделенное соответствие большей мощности $\theta 1$ -ритма у более молодых лиц (см. кластер 3 для $\theta 1$ в табл. 2) согласуется с эффектом возрастного снижения мощности низкочастотных ритмов ЭЭГ [16, 17], а меньшие значения ККС при наиболее низком уровне $\alpha 1$ колебаний (кластер 2 для $\alpha 1$ в табл. 2) — представлениям о сопряженной возрастной активации коры головного мозга на этих частотах и когнитивного дефицита [17–19]. Полученная в настоящем исследовании связь показателя ККС и фактора СТАБИЛЬНОСТЬ в $\beta 2$ -диапазоне может рассматриваться как отражение гиперактивации коры компенсаторного характера у пациентов с сердечно-сосудистыми заболеваниями. Ранее показано, что сочетанное уменьшение α -активности и рост мощности β -ритмов мо-

Таблица 2

Результаты ANOVA для мощности $\theta 1$ -, $\alpha 1$ - и $\beta 2$ -ритмов при сравнении выделенных кластеров

Переменная	$F(2,37)$	<i>P</i>	Кластер		
			1	2	3
$\theta 1$ -диапазон					
Мощность д/о	63,477	<0,0001	0,248	0,024	0,505
Возраст	5,132	0,011	58,87	57,29	53,09*
Мощность п/о	80,86	<0,0001	0,298	0,102	0,655
$\alpha 1$ -диапазон					
Мощность д/о	102,34	<0,0001	1,452	0,293	0,920
ККС	2,69	0,08	0,573	0,455*	0,575
Мощность п/о	89,79	<0,0001	1,430	0,312	0,894
$\beta 2$ -диапазон					
Мощность д/о	54,49	<0,0001	-0,278	-0,606	-0,468
Мощность п/о	73,24	<0,0001	-0,261	-0,651	-0,442
<p><i>Примечание:</i> д/о — до операции, п/о — после нее; * — $p < 0,01$ по сравнению с другими кластерами; остальные различия мощности между кластерами при $p < 0,0001$; курсивом выделены эффекты, характеризующие возможное функциональное значение кластеров; $F(2,37)$ — значения <i>F</i>-критерия и степени свободы в выполненном дисперсионном анализе.</p>					

жет свидетельствовать о повреждении региональных нейронных взаимодействий [20]. Есть предположение, что у пациентов с низким когнитивным статусом вследствие длительно существующего состояния хронической ишемии головного мозга перестройка электрической мозговой активности связана с "замедлением" ЭЭГ, т. е. преобладанием в электрической активности медленных ритмов, а не быстрых, как у тех, у кого более сохранен когнитивный статус [21]. Однако для анализа информационного значения координации разных ритмов ЭЭГ и их локального представительства для сохранности когнитивных функций у пациентов, перенесших коронарное шунтирование, требуются дальнейшие исследования.

Заключение

Результаты вычислительных экспериментов с кластеризацией данных нейрофизиологического тестирования с использованием трех представленных в настоящей статье формальных постановок выявили наивысшую эффективность задачи (7)–(11), (15) с использованием минимаксного критерия и адаптированного под нее алгоритма бинарных отсечений и ветвлений.

Время счета персонального компьютера до нахождения точных решений реализаций задачи кластеризации для показателей мощности ритмов ЭЭГ у 40 пациентов, при разных модификациях исходных данных и используемого критерия, находится в интервале от 15 мин до 2 ч. Приближенные решения могут быть получены за значительно более короткое время, что определяет хорошие перспективы применения инструментария с существенным увеличением размерностей реализаций задач.

Список литературы

1. Айвазян С.А., Бухштабер В.М., Еников И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989, 608 с.
2. Кельманов А. В., Пяткин А. В. NP-Трудность некоторых евклидовых задач разбиения конечного множества точек // Журнал вычислительной математики и математической физики. 2018. Т. 58, № 5. С. 852–856.
3. Кельманов А. В., Пяткин А. В., Хандеев В. И. О сложности некоторых максиминных задач кластеризации // Труды института математики и механики УрО РАН. 2018. Т. 24, № 4. С. 189–198.
4. Мезенцев Ю. А., Эстрайх И. В. Задачи и алгоритмы оптимизации расписаний параллельно-последовательных систем с неопределенными маршрутами обслуживания //

Доклады академии наук высшей школы РФ. Новосибирск: Изд-во НГТУ 2016. № 3 (32) С. 83–97.

5. Mezentsev Y., Estraykh I. Problems and optimization algorithms of schedules of parallel-serial systems with undefined service routes Constructive Nonsmooth Analysis and Related Topics (Dedicated to the Memory of V. F. Demyanov), CNSA 2017 <https://doi.org/10.1109/cnsa.2017.7973988>.

6. Vazirani V. V. Approximation algorithms. Springer, 2001, 378 p.

7. Pinedo M. Scheduling Theory, Algorithms, and Systems. Springer, 2008. 672 p.

8. Avdeenko T. V., Mesentsev Y. A. Efficient approaches to scheduling for unrelated parallel machines with release dates // IFAC-Papers Online (IFAC Proceedings Volumes). 2016. Vol. 49, Iss. 12: 8 IFAC conference on manufacturing modelling, management and control MIM 2016, France, Troyes, 28–30 June 2016. P. 1743–1748. URL: <http://www.sciencedirect.com/science/article/pii/S240589631631117X>.

9. Mezentsev Y. Binary cut-and-branch method for solving mixed integer programming problems // Constructive Nonsmooth Analysis and Related Topics (Dedicated to the Memory of V. F. Demyanov), CNSA 2017 <https://doi.org/10.1109/cnsa.2017.7973989>.

10. Мезенцев Ю. А. Метод бинарных отсечений и ветвлений целочисленного программирования // Доклады академии наук высшей школы РФ. Новосибирск: Изд-во НГТУ 2011. № 1(16). С. 12–25.

11. Тарасова И. В., Трубникова О. А., Кухарева И. Н., Барбараш О. Л., Барбараш Л. С. Годовая динамика нейрофизиологических показателей у пациентов, перенесших коронарное шунтирование с искусственным кровообращением // Комплексные проблемы сердечно-сосудистых заболеваний. 2015. № 1. С. 18–24.

12. Тарасова И. В., Трубникова О. А., Барбараш О. Л., Барбараш Л. С. Изменения электроэнцефалограммы у пациентов с ранней и стойкой послеоперационной когнитивной дисфункцией при коронарном шунтировании с искусственным кровообращением // Неврологический журнал. 2017. Т. 22, № 3. С. 136–141.

13. Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis // Brain Research Reviews. 1999. Vol. 29, N. 2–3. P. 169–195.

14. Bressler S. L., Richter C. G. Interareal oscillatory synchronization in top-down neocortical processing // *Curr. Opin. Neurobiol.* 2015. Vol. 31. P. 62–66. doi:10.1016/j.conb.2014.08.010.

15. Трубникова О. А., Каган Е. С., Куприянова Т. В., Малева О. В., Аргунова Ю. А., Кухарева И. Н. Нейропсихологический статус пациентов со стабильной ишемической болезнью сердца и факторы на него влияющие // Комплексные проблемы сердечно-сосудистых заболеваний. 2017. Т. 6, № 1. С. 112–121.

16. Cummins T. D., Finnigan S. Theta power is reduced in healthy cognitive aging // *Int J Psychophysiol.* 2007. Vol. 66, N. 1. P. 10–17.

17. Klass D. W., Brenner R. P. Electroencephalography of the elderly // *J Clin Neurophysiol.* 1995. Vol. 12, N. 2. P. 116–131.

18. Разумникова О. М. Закономерности старения мозга и способы активации его компенсаторных ресурсов // Успехи физиол. наук. 2015. Т. 46, № 2. С. 3–16.

19. Белоусова Л. В., Разумникова О. М., Вольф Н. В. Возрастные особенности связи интеллекта и характеристик ЭЭГ // Журн. Высш. Нервн. Деят. 2015. Т. 65, № 6. С. 699–705.

20. Zappasodi F., Olejarczyk E., Marzetti L., Assenza G., Pizzella V., Tecchio F. Fractal dimension of EEG activity senses neuronal impairment in acute stroke // *PLoS One.* 2014. 9(6): e100199. doi: 10.1371/journal.pone.0100199.

21. Тарасова И. В., Трубникова О. А., Барбараш О. Л., Барбараш Л. С. Диагностическое значение показателей электроэнцефалографии при ранней послеоперационной когнитивной дисфункции // Креативная кардиология. 2016. Т. 10, № 3. С. 220–230.

Yu. A. Mezentsev, Doctor of Technical Sciences, Professor, e-mail: mesyan@yandex.ru,
O. M. Razumnikova, Doctor of Biological Sciences, Professor, e-mail: razoum@mail.ru,
Novosibirsk State Technical University, Novosibirsk, Russia,
I. V. Tarasova, Doctor of Medical Sciences, Leading Researcher, e-mail: iriz78@mail.ru,
O. A. Trubnikova, Doctor of Medical Sciences, Head of Laboratory, e-mail: olgalet17@mail.ru,
Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russia

On some Problems of Big data Clustering by Minimax and Additive Criteria, Application in Medicine and Neurophysiology

The NP-hard clustering problem as applied to the data of neurophysiological studies (indicators of postoperative cognitive dysfunction) is considered. Variants of the problem of clustering in the form of mixed integer programming, including the use of continuous relaxation, reducing the complexity of solutions without loss of accuracy are given. The results of computational experiments on real data using the software implementation of the algorithm of binary cuts and branch are presented. They demonstrate the high efficiency of the developed toolkit.

Keywords: clustering, minimax quality criterion, additive criterion, linear relaxation, binary cutoff and branching algorithm, detection of postoperative cognitive dysfunction

Acknowledgements: This work was supported by the Russian Foundation for Basic Research, project no. 19-29-01017.

DOI: 10.17587/it.25.602-608

References

1. Ayzvyan S. A., Buchstaber V. M., Enikov I. S., Meshalkin L. D. Applied statistics: classification and reduction of dimension, Moscow, Finance and Statistics, 1989, 608 p. (in Russian).
2. Kelmanov A. V., Pyatkin A. V. NP-Difficulty of some Euclidean problems of partitioning a finite set of points, *Journal of Computational Mathematics and Mathematical Physics*, 2018, vol. 58, no. 5, pp. 852–856. (in Russian).
3. Kelmanov A. V., Pyatkin A. V., Khandeev V. I. On the complexity of some maximin clustering problems, *Proceedings of the Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences*, 2018, vol. 24, no. 4, pp. 189–198 (in Russian).
4. Mezentsev Y., Estraykh I. Tasks and algorithms for optimizing the schedules of parallel-sequential systems with undefined service routes, *Reports of the Academy of Sciences of Higher School of the Russian Federation*. Novosibirsk, NSTU Publ., 2016, no. 3 (32), pp. 83–97 (in Russian).
5. Mezentsev Y., Estraykh I. Problems and optimization algorithms of schedules of parallel-serial systems with undefined service routes *Constructive Nonsmooth Analysis and Related Topics* (Dedicated to the Memory of V. F. Demyanov), CNSA 2017, available at: <https://doi.org/10.1109/cnsa.2017.7973988>
6. Vazirani V. V. Approximation algorithms, Springer, 2001, 378 p.
7. Pinedo M. Scheduling Theory, Algorithms, and Systems (3rd. ed.), Springer, 2008. 672 p.
8. Avdeenko T. V., Mesentsev Y. A. Efficient approaches to scheduling for unrelated parallel machines with release dates, *IFAC-Papers Online (IFAC Proceedings Volumes)*, 2016, 49 (12), 8 IFAC conference on manufacturing modelling, management and control MIM, pp. 1743–1748.
9. Mezentsev Y. Binary cut-and-branch method for solving mixed integer programming problems, *Constructive Nonsmooth Analysis and Related Topics* (Dedicated to the Memory of V. F. Demyanov), CNSA, 2017 <https://doi.org/10.1109/cnsa.2017.7973989>
10. Mezentsev Y. Метод бинарных отсечений и ветвлений целочисленного программирования, *Reports of the Academy of Sciences of Higher School of the Russian Federation*, Novosibirsk, NSTU Publ., 2011, no. 1(16), pp. 12–25 (in Russian).
11. Tarasova I. V., Trubnikova O. A., Kukhareva I. N., Barbarash O. L., Barbarash L. S. Annual dynamics of neurophysiological parameters in patients undergoing coronary artery bypass surgery with artificial circulation, *Complex problems of cardiovascular diseases*, 2015, no. 1, pp. 18–24 (in Russian).
12. Tarasova I. V., Trubnikova O. A., Barbarash O. L., Barbarash L. S. Electroencephalogram changes in patients with early and resistant postoperative cognitive dysfunction during coronary artery bypass surgery with artificial blood circulation, *Neurological Journal*, 2017, vol. 22, no. 3, pp. 136–141 (in Russian).
13. Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis, *Brain Research Reviews*, 1999, 29, pp. 169–195.
14. Bressler S. L., Richter C. G. Interareal oscillatory synchronization in top-down neocortical processing, *Curr. Opin. Neurobiol.*, 2015, vol. 31, pp. 62–66, doi:10.1016/j.conb.2014.08.010.
15. Trubnikova O. A., Kagan E. S., Kupriyanova T. V., Mal'eva O. V., Argunova Yu. A., Kukhareva I. N. Neuropsychological status of patients with stable ischemic heart disease and factors influencing it, *Complex problems of cardiovascular diseases*, 2017, vol. 6, no. 1, pp. 112–121 (in Russian).
16. Cummins T. D., Finnigan S. Theta power is reduced in healthy cognitive aging, *Int J Psychophysiol.*, 2007, vol. 66, no. 1, pp. 10–17.
17. Klass D. W., Brenner R. P. Electroencephalography of the elderly, *J. Clin Neurophysiol.*, 1995, vol. 12, no. 2, pp. 116–131.
18. Razumnikova O. M. Patterns of aging of the brain and how to activate its compensatory resources, *Advances physiol. Science*, 2015, vol. 46, no. 2, pp. 3–16 (in Russian).
19. Belousova L. V., Razumnikova O. M., Volf N. V. Age features of communication of intelligence and EEG characteristics, *J. Higher Nervous Activity*, 2015, vol. 65, no. 6, pp. 699–705 (in Russian).
20. Zappasodi F., Olejarczyk E., Marzetti L., Assenza G., Pizzella V., Tecchio F. Fractal dimension of EEG activity senses neuronal impairment in acute stroke, *PLoS One*, 2014, vol. 9, no. 6: e100199, doi: 10.1371/journal.pone.0100199.
21. Tarasova I. V., Trubnikova O. A., Barbarash O. L., Barbarash L. S. Diagnostic value of electroencephalography indicators for early postoperative cognitive dysfunction, *Creative Cardiology*, 2016, vol. 10, no. 3, pp. 220–230 (in Russian).