На правах рукописи

Хандеев Владимир Ильич

АЛГОРИТМЫ С ОЦЕНКАМИ КАЧЕСТВА ДЛЯ КВАДРАТИЧНЫХ ЗАДАЧ КЛАСТЕРИЗАЦИИ С ФИКСИРОВАННЫМ ЦЕНТРОМ ОДНОГО ИЗ КЛАСТЕРОВ

01.01.09 — дискретная математика и математическая кибернетика

А в т о р е ф е р а т
диссертации на соискание учёной степени
кандидата физико-математических наук

Новосибирск — 2017

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте математики им. С.Л. Соболева Сибирского отделения Российской академии наук.

Научный руководитель:
Кельманов Александр Васильевич, доктор физико-математических наук, заведующий лабораторией.

Официальные оппоненты:
Хачай Михаил Юрьевич, доктор физико-математических наук, зав. отделом, Федеральное государственное бюджетное учреждение науки Институт математики и механики им. Н.Н. Красовского Уральского отделения Российской академии наук;

Симанчёв Руслан Юрьевич, кандидат физико-математических наук, зав. кафедрой, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Омский государственный университет им. Ф.М. Достоевского.

Ведущая организация: Федеральный исследовательский центр “Информатика и управление” Российской академии наук.

Защита состоится 13 сентября 2017 г. в 15 час. 00 мин. на заседании диссертационного совета Д 003.015.01 при Федеральном государственном бюджетном учреждении науки Институте математики им. С.Л. Соболева Сибирского отделения Российской академии наук по адресу: 630090, г. Новосибирск, пр. Академика Коптюга, 4.

С диссертацией можно ознакомиться в библиотеке Института математики им. С.Л. Соболева Сибирского отделения Российской академии наук и на сайте math.nsc.ru.

Автореферат разослан ”____” ____________ 2017 г.

Учёный секретарь
диссертационного совета
д.ф.-м.н. Ю.В. Шамардин

Общая характеристика работы

Актуальность работы[1]. Объект исследования работы — проблемы оптимизации. Предмет исследования — труднорешаемые экстремальные задачи разбиения множества и последовательности точек евклидова пространства. Цель исследования — изучение вопросов алгоритмической аппроксимируемости этих задач.

Рассматриваемые задачи в постановочном плане близки к классической NP-трудной в сильном смысле задаче MSSC (Minimum Sum-of-Squares Clustering). В этой задаче требуется разбить конечное множество точек евклидова пространства на несколько кластеров по критерию минимума суммы по всем кластерам внутрикластерных сумм квадратов расстояний от элементов кластеров до их геометрических центров (центроидов). Другое название задачи — k-means (k-средних). Основное отличие рассматриваемых NP-трудных в сильном смысле задач состоит в том, что для одного из кластеров внутрикластерная сумма равна сумме квадратов расстояний от элементов кластера до фиксированной точки — центра (без ограничения общности этой точкой может служить начало координат). В задачах кластеризации последовательности имеется дополнительное отличие — ограничения на элементы, входящие в кластеры.

Исследование мотивировано слабой изученностью задач и их актуальностью для анализа данных, распознавания образов, аппроксимации, компьютерной геометрии, статистики, а также для естественнонаучных и технических приложений, в которых требуется обработка и классификация данных численных экспериментов или результатов наблюдения за состояниями каких-либо объектов.

Цель работы — построение эффективных алгоритмов с гарантированными оценками качества (точности, трудоёмкости, вероятности несрабатывания) для модифицированной задачи MSSC, в которой центр одного из кластеров фиксирован, и для вариантов модифицированной задачи, ориентированных на разбиение последовательности при дополнительных ограничениях на элементы, входящие в кластеры.

Основные результаты.

1. Для квадратичной задачи разбиения конечного множества точек евклидова пространства на два кластера при фиксированном центре одного из кластеров:

[1]Работа поддержана грантами РФФИ 12-01-33028 мол-а-вед, 12-01-00090-а, 13-07-00070-а, 15-01-00462-а, 16-31-00186 мол-а, 16-07-00168-а, РНФ 16-11-10041.

(а) построен 2-приближённый полиномиальный алгоритм для случая задачи без ограничений на мощности кластеров;

(б) предложен рандомизированный алгоритм (ориентированный на случай задачи с ограничениями на мощности кластеров), который при заданных относительной ошибке и вероятности несрабатывания для установленных значений параметров находит приближённое решение за линейное время; найдены условия, при которых алгоритм полиномиален и асимптотически точен.

2. Для квадратичных евклидовых задач 2-кластеризации конечных множества и последовательности (с ограничениями на выбор элементов, входящих в кластеры) при фиксированном центре одного из кластеров и дополнительном ограничении на мощности кластеров:

(а) построены точные алгоритмы для случая целочисленных входов задачи; при фиксированной размерности пространства алгоритмы псевдополиномиальны;

(б) показано, что для общих случаев задач не существует полностью полиномиальных приближённых схем (FPTAS), если $P \neq NP$; такие схемы построены для случаев задач, в которых размерность пространства фиксирована.

3. Для квадратичной евклидовой задачи многокластерного разбиения конечной последовательности точек с ограничениями на выбор внутрикластерных элементов при фиксированном центре одного из кластеров построены 2-приближённые алгоритмы, ориентированные как на случай задачи без ограничений на мощности кластеров, так и на случай с ограничениями; алгоритмы полиномиальны при фиксированном числе кластеров.

Научная новизна. Все результаты диссертации снабжены строгими доказательствами, а их новизну раскрывают следующие аргументы (по каждому основному результату).

1. Алгоритм из п. 1 (а) имеет меньшую трудоёмкость по сравнению с лучшим из известных алгоритмов при той же, как и у этого алгоритма, точности. Алгоритм из п. 1 (б) является первым алгоритмом рандомизированного типа, предложенным для задачи из этого пункта.

2. Алгоритм кластеризации множества из п. 2 (а) является новым решением задачи, послужившим важным промежуточным результатом, на котором основана идея построения оригинальных аппроксимационных схем из п. 2 (б). Алгоритм разбиения последовательности из п. 2 (а) построен впервые. Факт несуществования схемы FPTAS для общих случаев задач из п. 2 также установлен впервые; результаты по построению приближённых схем для указанных в п. 2 (б) случаев задач

приоритетны.

3. На настоящее время результаты п. 3 являются единственными алгоритмами с гарантированными оценками точности, предложенными для задач из этого пункта.

Методы исследований. В работе используются методы дискретной оптимизации, геометрии, теории вероятностей.

Теоретическая значимость и практическая ценность. Работа носит теоретический характер. Предложенные алгоритмы имеют теоретическую значимость для математических проблем анализа данных и распознавания образов, дискретной оптимизации, аппроксимации, компьютерной геометрии. Практическая ценность результатов (алгоритмов) состоит в том, что они могут быть использованы в естественнонаучных и технических приложениях для создания эффективных компьютерных технологий с теоретическими гарантиями качества, ориентированных, в частности, на обработку сигналов, изображений, результатов численных экспериментов, дистанционный мониторинг и др.

На защиту выносится совокупность эффективных алгоритмов с гарантированными оценками качества для решения NP-трудных в сильном смысле квадратичных задач кластеризации конечных множества и последовательности точек евклидова пространства с фиксированным центром одного из кластеров.

Личный вклад автора. Постановки задач предложены научным руководителем. Подходы к построению алгоритмов найдены совместно. Доказательства утверждений получены соискателем лично. Конфликт интересов с соавторами отсутствует.

Апробация работы. Результаты работы докладывались на семинарах ИМ СО РАН (часть из них отмечены в качестве важнейших), кафедры Теоретической кибернетики НГУ (часть их них отмечены Премией им. А.А. Ляпунова) и на следующих всероссийских и международных конференциях: «Проблемы оптимизации и экономические приложения» (Омск, 2012, 2015); «Intelligent Data Processing: Theory and Applications» (Черногория, 2012, Греция, 2014, Испания, 2016); «Discrete Optimization and Operations Research» (Новосибирск, 2013, Владивосток, 2016); «Математические методы распознавания образов» (Казань, 2013, Светлогорск, 2015); «Optimization and applications» (Черногория, 2013, 2014, 2015, 2016); «Methods of Optimization and Their Applications» (Иркутск, 2014); «European Chapter on Combinatorial Optimization» (Италия, 2015); «Математическое программирование и приложения» (Екатеринбург, 2015).

Публикации. По теме диссертации опубликовано 29 работ, из них

20 — тезисы докладов, 9 работ — в изданиях, входящих в список ВАК, в том числе 5 — в журналах, индексируемых системой цитирования Web of Science, 9 — Scopus, 9 — RSCI (ядро РИНЦ).

Структура и объем диссертации. Работа состоит из введения, трёх глав, заключения и списка литературы. Объём диссертации — 111 страниц. Список литературы содержит 88 источников.

Содержание работы

Во **введении** обоснована актуальность работы, приведено краткое изложение содержания работы, сформулированы основные результаты и раскрыта их новизна.

В **первой главе** представлены истоки задач, их трактовки и приложения, а также результаты для классической задачи MSSC, близкой в постановочном плане к рассматриваемым задачам. Своими корнями эта известная около 60 лет задача уходит к классическим работам Фишера. С ней связаны тысячи публикаций математического и прикладного плана. Основные математические результаты отмечены в обзоре.

Исследуемая двухкластерная модификация этой задачи с фиксированным (в нуле) центром одного кластера индуцируется[2] одной из проблем поиска максимально правдоподобного повторяющегося фрагмента в числовой последовательности на фоне гауссовского шума (с нулевым средним), и связана с проблемами помехоустойчивого обнаружения и распознавания квазипериодически повторяющихся импульсов[3,4]. Сильная NP-трудность задачи и её варианта с ограничениями на мощность следует из сильной NP-трудности полиномиально эквивалентной задачи на максимум и её варианта с мощностными ограничени-

[2]Кельманов А.В., Хамидуллин С.А, Кельманова М.А. Совместное обнаружение и оценивание повторяющегося фрагмента в зашумленной числовой последовательности при заданном числе квазипериодических повторов // Тез. докл. Российской конф. «Дискретный анализ и исследование операций» (DAOR-4). Новосибирск, 2004. — С. 185.

[3]Kel'manov A.V., Jeon B. A Posteriori Joint Detection and Discrimination of Pulses in a Quasiperiodic Pulse Train // IEEE Trans. Signal Processing. — 2004. — Vol. 52, № 3. — P. 645–656.

[4]Кельманов А. В., Хамидуллин С. А. Распознавание квазипериодической последовательности, образованной из заданного числа одинаковых подпоследовательностей // Сиб. журн. индустр. математики. — 1999. — Т.2, № 1. — С. 53–74.

ями[5,6,7,8,9], которые также индуцируются отмеченной выше прикладной проблемой оптимального обнаружения. Сильная NP-трудность исследуемых родственных задач разбиения последовательности на два и более кластеров при фиксированном в нуле центре одного из кластеров установлена[10] лишь несколько лет назад. Обзор немногочисленных алгоритмических результатов по исследованиям рассматриваемых задач приводится после их формулировок в следующих главах. Слабая изученность задач определила направление исследований и характер полученных результатов.

Во **второй главе** рассматриваются задачи 2-кластеризации конечного множества точек евклидова пространства при фиксированном (в начале координат) центре одного из кластеров.

Задача 1. *Дано:* N-элементное множество $\mathcal{Y}$ точек из $\mathbb{R}^d$. *Найти:* разбиение $\mathcal{Y}$ на кластеры $\mathcal{C}$ и $\mathcal{Y} \setminus \mathcal{C}$ такое, что

$$S(\mathcal{C}) = \sum_{y \in \mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 + \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y\|^2 \to \min, \tag{1}$$

где $\overline{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ — центроид подмножества $\mathcal{C}$.

Ранее для этой задачи был известен[11] 2-приближённый алгоритм с трудоёмкостью $\mathcal{O}(dN^3)$ и точный алгоритм[12] с трудоёмкостью $\mathcal{O}(d^2 N^{2d})$.

Для этой задачи в работе обоснован (п. 1 основных результатов)

---

[5] Кельманов А.В., Пяткин А.В. О сложности одного из вариантов задачи выбора подмножества «похожих» векторов // Доклады РАН. — 2008. — Т. 421, № 5. — С. 590–592.

[6] Кельманов А.В., Пяткин А.В. Об одном варианте задачи выбора подмножества векторов // Дискрет. анализ и исслед. опер. — 2008. — Т. 15, № 5. — С. 20–34.

[7] Гимади Э.Х., Кельманов А.В., Кельманова М.А., Хамидуллин С.А. Апостериорное обнаружение в числовой последовательности квазипериодического фрагмента при заданном числе повторов // Сиб. журн. индустр. математики. — 2006. — Т. 9, № 1(25). — С. 55–74.

[8] E.Kh. Gimadi, A.V. Kel'manov, M.A. Kel'manova, S.A. Khamidullin. A posteriori detecting a quasiperiodic fragment in a numerical sequence // Pattern Recognition and Image Analysis. — 2008. — Vol. 18, № 1. — P. 30–42.

[9] Бабурин А. Е., Гимади Э. Х., Глебов Н. И., Пяткин А. В. Задача отыскания подмножества векторов с максимальным суммарным весом // Дискрет. анализ и исслед. опер. Серия 2. — 2007. — Т. 14, № 1. — С. 32–42.

[10] Кельманов А. В., Пяткин А. В. О сложности некоторых задач кластерного анализа векторных последовательностей // Дискрет. анализ и исслед. опер. — 2013. — Т. 20, № 2. — С. 47–57.

[11] Долгушев А.В., Кельманов А.В. Приближенный алгоритм решения одной задачи кластерного анализа // Дискрет. анализ и исслед. опер. — 2011. — Т. 18, № 2. — С. 29–40.

[12] Гимади Э.Х., Пяткин А.В., Рыков И.А. О полиномиальной разрешимости некоторых задач выбора подмножеств векторов в евклидовом пространстве фик-

Алгоритм[13] $\mathcal{A}_1$(2-приближённый полиномиальный алгоритм).

Вход алгоритма: $\mathcal{Y}$. `Шаг 1.` Для каждого $x \in \mathcal{Y}$ найдём множество $\mathcal{B}(x) = \{y \in \mathcal{Y} \mid 2(y, x) > \|x\|^2\}$ и вычислим значение функции

$$Q(\mathcal{C}, x) = \sum_{y \in \mathcal{C}} \|y - x\|^2 + \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y\|^2$$

при $\mathcal{C} = \mathcal{B}(x)$. `Шаг 2.` В семействе $\{\mathcal{B}(x) \mid x \in \mathcal{Y}\}$, найденном на шаге 1, выберем в качестве решения $\mathcal{C}_A$ то подмножество $\mathcal{B}(x)$, для которого значение $Q(\mathcal{B}(x), x)$ минимально. Если минимальному значению соответствует несколько подмножеств, то выберем любое из них. *Выход алгоритма:* $\mathcal{C}_A$.

Теорема 2.1. Алгоритм $\mathcal{A}_1$ находит 2-приближённое решение задачи 1 за время $\mathcal{O}(dN^2)$. Оценка 2 точности алгоритма достижима.

Задача 2. *Дано:* N-элементное множество $\mathcal{Y}$ точек из $\mathbb{R}^d$ и натуральное число M. *Найти:* разбиение $\mathcal{Y}$ на кластеры $\mathcal{C}$ и $\mathcal{Y} \setminus \mathcal{C}$ такое, что целевая функция (1) минимальна, при ограничении $|\mathcal{C}| = M$.

Для задачи 2 были известны: (а) точный алгоритм[12] с трудоёмкостью $\mathcal{O}(d^2 N^{2d})$; (б) точные алгоритмы[14,15]для случая целочисленных входов с трудоёмкостью $\mathcal{O}(Nd^{d+1}(MD)^{d-1})$ и $\mathcal{O}(dMN(2MD)^{d-1})$, где D — максимальное абсолютное значение компонент входных точек; (в) 2-приближённый полиномиальный алгоритм[11] с трудоёмкостью $\mathcal{O}(dN^2)$; (г) схема PTAS[16] с трудоёмкостью $\mathcal{O}(dN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$, где ε — относительная погрешность.

Для подкласса задачи 2, в котором $\mathcal{Y} \subset \mathbb{Z}^d$, а размерность d пространства фиксирована, в работе предложен (п. 2 (а) результатов)

Алгоритм $\mathcal{A}_2$ (точный псевдополиномиальный алгоритм).

Вход алгоритма: $\mathcal{Y}$, M. `Шаг 1.` Для каждой точки $x \in \mathcal{G}$, где $\mathcal{G}$ — многомерная решётка кубической формы размера $2D$ с расстоянием

сированной размерности // Дискрет. анализ и исслед. опер. — 2008. — Т. 15, № 6. — С. 11–19.

[13]В автореферате все алгоритмы записаны укрупнёнными шагами; детальная запись представлена в диссертации.

[14]Бабурин А.Е., Гимади Э.Х., Глебов Н.И., Пяткин А.В. Задача отыскания подмножества векторов с максимальным суммарным весом // Дискрет. анализ и исслед. опер. — 2007. — Т. 14, № 1. — С. 32–42.

[15]Гимади Э.Х., Глазков Ю.В., Рыков И.А. О двух задачах выбора подмножества векторов с целочисленными координатами в евклидовом пространстве с максимальной нормой суммы размерности // Дискрет. анализ и исслед. опер. — 2008. — Т. 15, № 4. — С. 30—43.

[16]Долгушев А.В., Кельманов А.В., Шенмайер В.В. Полиномиальная аппроксимационная схема для одной задачи разбиения конечного множества на два кластера // Тр. Ин-та математики и механики УрО РАН. — 2015. — Т. 21, № 3. — С. 100–109.

$\frac{1}{M}$ между узлами и центром в начале координат, а D — максимальное абсолютное значение координат точек множества $\mathcal{Y}$, построим подмножество $\mathcal{B}_M(x)$ — допустимое решение задачи — из M элементов $\mathcal{Y}$, имеющих наибольшие проекции на направление, задаваемое точкой x. Вычислим значение $Q(\mathcal{B}_M(x), x)$. `Шаг 2`. Найдём точку $x_A = \arg\min_{x \in \mathcal{G}} Q(\mathcal{B}_M(x), x)$ (если минимальному значению соответствует несколько точек, то выберем любую из них) и соответствующее ей подмножество $\mathcal{B}_M(x_A)$. В качестве решения задачи возьмём подмножество $\mathcal{C}_A = \mathcal{B}_M(x_A)$. *Выход алгоритма:* $\mathcal{C}_A$.

Теорема 2.2. Пусть в условиях задачи 2 точки из множества $\mathcal{Y}$ имеют целочисленные координаты из интервала $[-D, D]$. Тогда алгоритм $\mathcal{A}_2$ находит оптимальное решение за время $\mathcal{O}(dN(2MD+1)^d)$.

В случае фиксированной размерности d пространства алгоритм псевдополиномиален, так как в этом случае трудоёмкость алгоритма оценивается величиной $\mathcal{O}(N(MD)^d)$.

Задача 2 NP-трудна в сильном смысле, но имеет числовые входы. Поэтому рассматривался вопрос о существовании схемы FPTAS. Доказана (п. 2 (б) основных результатов)

Теорема 2.3. Если P $\neq$ NP, то для задачи 2 не существует схемы FPTAS.

Для задачи 2 предложен (п. 2 (б) основных результатов)

Алгоритм $\mathcal{A}_3$ (аппроксимационная схема).

Вход алгоритма: $\mathcal{Y}$, M и число $\varepsilon > 0$. Для каждой точки $y \in \mathcal{Y}$ выполним шаги 1–5. `Шаг 1.` Построим подмножество $\mathcal{B}_M(y)$ из M элементов $\mathcal{Y}$, имеющих наибольшие проекции на направление, задаваемое точкой y. `Шаг 2.` Вычислим значения $S(\mathcal{B}_M(y))$, $h(y, \varepsilon) = \sqrt{\frac{2\varepsilon}{dM} S(\mathcal{B}_M(y))}$ и $H(y) = \sqrt{\frac{1}{M} S(\mathcal{B}_M(y))}$. `Шаг 3.` Если $S(\mathcal{B}_M(y)) = 0$, то подмножество $\mathcal{B}_M(y)$ объявим результатом $\mathcal{C}_A$ работы алгоритма; выход. Иначе переходим к следующему шагу. `Шаг 4.` Построим многомерную решётку $\mathcal{G}(y, h, H + h/2)$ кубической формы размера $2H + h$ с расстоянием h между узлами с центром в точке y. `Шаг 5.` Для каждого узла x решётки $\mathcal{G}(y, h, H+h/2)$ найдём подмножество $\mathcal{B}_M(x)$ — допустимое решение задачи — из M элементов $\mathcal{Y}$, имеющих наибольшие проекции на направление, задаваемое точкой x, и вычислим значение $S(\mathcal{B}_M(x))$. `Шаг 6.` В семействе $\{\mathcal{B}_M(x) \mid x \in \mathcal{G}(y, h, H + h/2),\ y \in \mathcal{Y}\}$ выберем в качестве решения $\mathcal{C}_A$ то подмножество $\mathcal{B}_M(x)$, для которого значение $S(\mathcal{B}_M(x))$ минимально. Если оптимуму соответствует несколько подмножеств, то выберем любое из них. *Выход алгоритма:* $\mathcal{C}_A$.

Теорема 2.4. Для любого $\varepsilon > 0$ алгоритм $\mathcal{A}_3$ находит $(1+\varepsilon)$-приближённое решение задачи 2 за время[17] $\mathcal{O}(dN^2(\sqrt{\frac{2d}{\varepsilon}}+2)^d)$.

При условии фиксированной размерности d пространства алгоритм реализует схему FPTAS, т.к. при этом условии время работы алгоритма есть величина $\mathcal{O}(N^2(1/\varepsilon)^{d/2})$.

Наконец, в главе 2 для задачи 2 предложен (п. 2 (в) основных результатов)

Алгоритм $\mathcal{A}_4$ (рандомизированный алгоритм).

Вход алгоритма: $\mathcal{Y}$, M, натуральный параметр k. `Шаг 1`. Сформируем мультимножество $\mathcal{T}$ из k независимых случайных выборок (с возвращением) по одному элементу из $\mathcal{Y}$. `Шаг 2`. Для каждого непустого $\mathcal{H} \subseteq \mathcal{T}$ вычислим центроид $\overline{y}(\mathcal{H})$ и сформируем подмножество $\mathcal{B}_M(\overline{y}(\mathcal{H}))$ — допустимое решение задачи — из M элементов $\mathcal{Y}$, имеющих наибольшие проекции на направление, задаваемое этим центроидом. Вычислим $S(\mathcal{B}_M(\overline{y}(\mathcal{H})))$. `Шаг 3`. В семействе подмножеств, найденных на шаге 2, выберем в качестве решения то подмножество $\mathcal{C}_A = \mathcal{B}_M(\overline{y}(\mathcal{H}))$, для которого значение $S(\mathcal{B}_M(\overline{y}(\mathcal{H})))$ минимально. Если оптимальному значению соответствует несколько подмножеств, то выберем любое из них. *Выход алгоритма:* $\mathcal{C}_A$.

Теорема 2.5. Для произвольных вещественного $\delta \in (0,1)$ и натурального $t \le k$ алгоритм $\mathcal{A}_4$ находит $(1+\frac{1}{\delta t})$-приближённое решение задачи 2 за время $\mathcal{O}(2^k d(k+N))$ с вероятностью не менее $1-(\delta+\alpha)$, где $\alpha = \sum_{i=0}^{t-1}\binom{k}{i}\left(\frac{M}{N}\right)^i\left(1-\frac{M}{N}\right)^{k-i}$.

Следствие 2.2. Допустим, что $M \ge \beta N$, где $\beta \in (0,1)$ — некоторая константа. Тогда для заданных $\varepsilon > 0$ и $\gamma \in (0,1)$ при $k = \max(\lceil \frac{2}{\beta}\lceil \frac{2}{\gamma\varepsilon}\rceil\rceil, \lceil \frac{8}{\beta}\ln\frac{2}{\gamma}\rceil)$ алгоритм $\mathcal{A}_4$ находит $(1+\varepsilon)$-приближённое решение задачи 2 с вероятностью не менее $1-\gamma$ за время $\mathcal{O}(dN)$.

Условия асимптотической точности алгоритма устанавливает

Теорема 2.6. Пусть в условиях теоремы 2.5 имеют место равенства $k = \lceil \log_2 N\rceil$, $\delta = (\log_2 N)^{-1/2}$, $t = \lceil \frac{kM}{2N}\rceil$. Тогда если $M \ge \beta N$, где $\beta \in (0,1)$ — некоторая константа, то алгоритм $\mathcal{A}_2$ находит $(1+\varepsilon_N)$-приближённое решение задачи 2 с вероятностью $1-\gamma_N$ за время $\mathcal{O}(dN^2)$, где $\varepsilon_N \le \frac{2}{\beta}(\log_2 N)^{-1/2}$, $\gamma_N \le (\log_2 N)^{-1/2} + N^{-\frac{\beta}{8\ln 2}}$.

В **третьей главе** рассматриваются задачи кластеризации конечной последовательности точек евклидова пространства при фиксированном

[17] При аккуратной организации вычислений формула трудоёмкости алгоритма именно такая, что слегка отличается от формулы $\mathcal{O}(dN^2(\sqrt{2d/\varepsilon}+1)^d)$, указанной в статье [4].

центре одного из кластеров и ограничениях на элементы, входящие в кластеры. По сути, результаты из этой главы развивают результаты главы 2. Учёт ограничений в задачах кластеризации последовательности потребовал применения техники (схем) динамического программирования.

Задача 3. *Дано:* последовательность $\mathcal{Y} = (y_1, \ldots, y_N)$ точек из $\mathbb{R}^d$, натуральные числа $T_{\min}$, $T_{\max}$ и $M > 1$. *Найти:* набор $\mathcal{M} = (n_1, \ldots, n_M)$, где $n_m \in \mathcal{N} = \{1, \ldots, N\}$, $m = 1, \ldots, M$, номеров элементов $\mathcal{Y}$ такой, что

$$F(\mathcal{M}) = \sum_{j \in \mathcal{M}} \|y_j - \overline{y}(\mathcal{M})\|^2 + \sum_{i \in \mathcal{N} \setminus \mathcal{M}} \|y_i\|^2 \to \min,$$

где $\overline{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$, при ограничениях

$$T_{\min} \le n_m - n_{m-1} \le T_{\max} \le N, \quad m = 2, \ldots, M, \tag{2}$$

на элементы набора $\mathcal{M}$.

К настоящему времени для задачи 3 был получен лишь один алгоритмический результат, а именно: 2-приближённый полиномиальный алгоритм[18] с трудоёмкостью $\mathcal{O}(N^2(MN + d))$.

Для случая задачи 3, в котором $\mathcal{Y} \subset \mathbb{Z}^d$, а размерность d пространства фиксирована, в работе предложен точный псевдополиномиальный алгоритм $\mathcal{A}_5$ (п. 2 (а) основных результатов). Пошаговая запись этого алгоритма аналогична записи алгоритма $\mathcal{A}_2$ (ориентированного на разбиение множества). Однако на шаге 1 алгоритма $\mathcal{A}_5$, в отличие от алгоритма $\mathcal{A}_2$, в качестве допустимого решения выбирается не множество $\mathcal{B}_M(x)$, а оптимальное решение $\mathcal{M}^x$ задачи

$$G^x(\mathcal{M}) = \sum_{n \in \mathcal{M}} (2\langle y_n, x\rangle - \|x\|^2) \to \max_{\mathcal{M} \subseteq \mathcal{N}} \tag{3}$$

при ограничениях (2) на элементы набора $\mathcal{M}$. Это решение — набор $\mathcal{M}^x$ — находится за время $\mathcal{O}(N(M(T_{\max} - T_{\min} + 1) + d))$ с помощью схемы[18] динамического программирования. На шаге 2 в качестве точки x_A выбирается та точка $x \in \mathcal{G}$, для которой значение $G^x(\mathcal{M}^x)$ максимально; выходом алгоритма является набор $\mathcal{M}^{x_A}$.

Теорема 3.1. Пусть в условиях задачи 3 точки последовательности $\mathcal{Y}$ имеют целочисленные координаты из интервала $[-D, D]$. Тогда

[18] Кельманов А. В., Хамидуллин С. А. Приближённый полиномиальный алгоритм для одной задачи разбиения последовательности // Дискрет. анализ и исслед. опер. — 2014. — Т. 21, № 1. — С. 53–66.

алгоритм $\mathcal{A}_5$ находит оптимальное решение за время $\mathcal{O}(N(M(T_{\max} - T_{\min} + 1) + d)(2MD + 1)^d)$.

При фиксированной размерности пространства алгоритм псевдополиномиален, так как при этом условии его трудоёмкость оценивается величиной $\mathcal{O}(N^3(MD)^d)$.

В этой же главе показано, что если P $\neq$ NP, то для задачи 3 не существует схемы FPTAS, и предложена приближённая схема — алгоритм $\mathcal{A}_6$ (п. 2 (б) основных результатов). Подход к построению этого алгоритма сходен с подходом, предложенным в главе 2 для алгоритма $\mathcal{A}_3$ разбиения множества. Основное отличие алгоритма $\mathcal{A}_6$ состоит в том, что допустимые решения — наборы номеров последовательности — находятся в результате решения задачи (3) с помощью схемы[18] динамического программирования.

Теорема 3.2. Для любого фиксированного $\varepsilon > 0$ алгоритм $\mathcal{A}_6$ находит $(1+\varepsilon)$-приближённое решение задачи 3 за время[19] $\mathcal{O}(N^2(M(T_{\max} - T_{\min} + 1) + d)(\sqrt{\frac{2d}{\varepsilon}} + 2)^d)$.

Если размерность d пространства ограничена константой, то трудоёмкость алгоритма оценивается величиной $\mathcal{O}(MN^3(1/\varepsilon)^{d/2})$ и он реализует схему FPTAS.

Кроме того, в этой главе для задачи 3 предложен рандомизированный алгоритм $\mathcal{A}_7$. Идея построения этого алгоритма аналогична идее построения алгоритма $\mathcal{A}_4$ разбиения множества, предложенного в главе 2. Ключевое отличие состоит в том, что семейство допустимых решений (наборов номеров элементов последовательности) находится, как и в алгоритмах $\mathcal{A}_5$ и $\mathcal{A}_6$, с использованием схемы динамического программирования.

Для алгоритма $\mathcal{A}_7$ доказаны теоремы 3.3, 3.4 и следствие 3.2 — аналоги теорем 2.5, 2.6 и следствия 2.2 для алгоритма $\mathcal{A}_4$. Эти утверждения показывают, что алгоритм $\mathcal{A}_7$ имеет те же вероятностные и аппроксимационные оценки, что и алгоритм $\mathcal{A}_4$, но увеличенное в MN раз время работы — $\mathcal{O}(dMN^2)$ в неасимптотическом случае и $\mathcal{O}(dMN^3)$ в асимптотическом.

Наконец, в главе 3 исследовались обобщения задач 1 и 2 — (далее) задачи 4 и 5, которые индуцируются прикладными проблемами анализа временных рядов. Для обеих исследуемых задач, за исключением их частных случаев, к настоящему времени отсутствовали какие-либо

[19]При аккуратной организации вычислений формула трудоёмкости алгоритма именно такая, что слегка отличается от формулы $\mathcal{O}(N^2(M(T_{\max} - T_{\min} + 1) + d)(\sqrt{2d/\varepsilon} + 1)^d)$, указанной в статье [6].

эффективные алгоритмы с оценками точности.

Задача 4. *Дано:* последовательность $\mathcal{Y} = (y_1, \ldots, y_N)$ точек из $\mathbb{R}^d$, натуральные числа $T_{\min}$, $T_{\max}$, L и M. *Найти:* непустые непересекающиеся наборы $\mathcal{M}_1, \ldots, \mathcal{M}_L$ номеров элементов $\mathcal{Y}$ такие, что

$$F(\mathcal{M}_1, \ldots, \mathcal{M}_L) = \sum_{l=1}^{L} \sum_{j \in \mathcal{M}_l} \|y_j - \overline{y}(\mathcal{M}_l)\|^2 + \sum_{i \in \mathcal{N} \setminus \mathcal{M}} \|y_i\|^2 \longrightarrow \min, \quad (4)$$

где $\mathcal{M} = \cup_{l=1}^{L} \mathcal{M}_l$, $\overline{y}(\mathcal{M}_l) = \frac{1}{|\mathcal{M}_l|} \sum_{j \in \mathcal{M}_l} y_j$ — центроид $\{y_j \mid j \in \mathcal{M}_l\}$, при ограничениях: (i) мощность объединённого набора $\mathcal{M}$ равна M, (ii) в последовательности, образованной конкатенацией наборов $\mathcal{M}_1, \ldots, \mathcal{M}_L$, номера упорядочены по возрастанию при условии, что элементы каждого набора образуют возрастающую последовательность, и (iii) номера из объединённого набора $\mathcal{M} = (n_1, \ldots, n_M)$ связаны неравенствами (2).

Для задачи 4 предложен (п. 3 основных результатов)

Алгоритм $\mathcal{A}_8$ (2-приближённый алгоритм).

Вход алгоритма: $\mathcal{Y}$, $T_{\min}$, $T_{\max}$, M и L. `Шаг 1`. Для каждого набора $x = (x_1, \ldots, x_L) \in \mathcal{Y}^L$ найдём решение $\{\mathcal{M}_1^x, \ldots, \mathcal{M}_L^x\}$ задачи

$$\sum_{l=1}^{L} \sum_{j \in \mathcal{M}_l} \left(2\langle y_j, x_l \rangle - \|x_l\|^2\right) \to \max \quad (5)$$

при тех же, что и в задаче 4, ограничениях на элементы $\{\mathcal{M}_1, \ldots, \mathcal{M}_L\}$. `Шаг 2`. Во множестве $\{\mathcal{M}_1^x, \ldots, \mathcal{M}_L^x \mid x \in \mathcal{Y}^L\}$ в качестве решения $\{\mathcal{M}_1^A, \ldots, \mathcal{M}_L^A\}$ выберем то семейство, для которого функция (5) имеет наибольшее значение. Если наибольшему значению соответствует несколько семейств, то выберем любое из них. *Выход алгоритма:* $\{\mathcal{M}_1^A, \ldots, \mathcal{M}_L^A\}$.

Конструктивно показано, что на шаге 1 при каждом фиксированном x задача (5) разрешима за время $\mathcal{O}(LN(M(T_{\max} - T_{\min} + 1) + d))$ с помощью алгоритма[20] динамического программирования.

Теорема 3.5. Алгоритм $\mathcal{A}_8$ находит 2-приближённое решение задачи 4 за время $\mathcal{O}(LN^{L+1}(M(T_{\max} - T_{\min} + 1) + d))$. Оценка 2 точности алгоритма достижима.

Поскольку $T_{\max} - T_{\min} + 1 \leq N$, время работы алгоритма оценивается величиной $\mathcal{O}(LN^{L+1}(MN + d))$; при фиксированном L алгоритм полиномиален.

---

[20] Кельманов А.В., Михайлова Л.В. Совместное обнаружение в квазипериодической последовательности заданного числа фрагментов из эталонного набора и ее разбиение на участки, включающие серии одинаковых фрагментов // Журн. вычисл. математики и мат. физики. — 2006. — Т. 46, № 1. — С. 172–189.

Задача 5. *Дано:* последовательность $\mathcal{Y} = (y_1, \ldots, y_N)$ точек из $\mathbb{R}^d$, натуральные числа $T_{\min}$, $T_{\max}$ и L. *Найти:* непустые непересекающиеся наборы $\mathcal{M}_1, \ldots, \mathcal{M}_L$ номеров элементов $\mathcal{Y}$ такие, что минимальна целевая функция (4), при ограничениях: (i) в последовательности, образованной конкатенацией наборов $\mathcal{M}_1, \ldots, \mathcal{M}_L$, номера упорядочены по возрастанию при условии, что элементы каждого набора образуют возрастающую последовательность, и (ii) номера из объединённого набора $\mathcal{M} = (n_1, \ldots, n_M)$ связаны неравенствами (2), где M зависит от оптимизируемых переменных.

Для задачи 5 предложен (п. 3 основных результатов)

Алгоритм $\mathcal{A}_9$ (2-приближённый алгоритм).

Вход алгоритма: $\mathcal{Y}$, $T_{\min}$, $T_{\max}$ и L. `Шаг 1`. Для каждого набора $x = (x_1, \ldots, x_L) \in \mathcal{Y}^L$ найдём решение $\{\mathcal{M}_1^x, \ldots, \mathcal{M}_L^x\}$ задачи максимизации функции (5) при тех же, что и в задаче 5, ограничениях на элементы семейства $\{\mathcal{M}_1, \ldots, \mathcal{M}_L\}$. `Шаг 2`. Среди решений, найденных на шаге 1, выберем то семейство $\{\mathcal{M}_1^A, \ldots, \mathcal{M}_L^A\}$, для которого значение $F(\mathcal{M}_1^A, \ldots, \mathcal{M}_L^A)$ минимально. Если минимальному значению соответствует несколько семейств, то выберем любое из них. *Выход алгоритма:* $\{\mathcal{M}_1^A, \ldots, \mathcal{M}_L^A\}$.

Установлено, что на шаге 1 при каждом фиксированном x решение задачи максимизации может быть найдено за время $\mathcal{O}(LN(T_{\max} - T_{\min} + d))$ с помощью алгоритма[21] динамического программирования.

Теорема 3.6. Алгоритм $\mathcal{A}_9$ находит 2-приближённое решение задачи 5 за время $\mathcal{O}(LN^{L+1}(T_{\max} - T_{\min} + d))$.

Поскольку $T_{\max} - T_{\min} + 1 \leq N$, время работы алгоритма оценивается величиной $\mathcal{O}(LN^{L+1}(N + d))$; при фиксированном L алгоритм полиномиален.

В **заключении** сформулированы основные результаты работы и обозначены перспективы дальнейших исследований.

Публикации автора по теме диссертации

Индексируемые в базах цитирования

[1] Кельманов А.В., Хандеев В.И. Полиномиальный алгоритм с оценкой точности 2 для решения одной задачи кластерного анализа //

[21] Кельманов А.В., Михайлова Л.В. Апостериорное обнаружение квазипериодических фрагментов из эталонного набора в числовой последовательности и ее разбиение на участки, включающие серии одинаковых фрагментов // Журн. вычисл. математики и мат. физики. — 2008. — Т. 48, № 5. — С. 899–915.

Дискрет. анализ и исслед. опер. — 2013. — Т. 20, № 4. — С. 36–45. RSCI (ядро РИНЦ).
Kel'manov A.V., Khandeev V.I. A 2-Approximation Polynomial Algorithm for a Clustering Problem // J. Appl. Ind. Math. — 2013. — V. 7, № 4. — P. 515–521. Scopus. DOI: 10.1134/S1990478913040066.

[2] Кельманов А.В., Хандеев В.И. Рандомизированный алгоритм для одной задачи двухкластерного разбиения множества векторов // Журн. вычисл. математики и мат. физики. — 2015. — Т. 55, № 2. — С. 335–344. RSCI (ядро РИНЦ). DOI: 10.7868/S0044466915020131.
Kel'manov A.V., Khandeev V.I.. A Randomized Algorithm for Two-Cluster Partition of a Set of Vectors // Comput. Math. Math. Phys. 2015. — V. 55, № 2. — P. 330–339. Scopus, WoS. DOI: 10.1134/S096554251502013X.

[3] Кельманов А.В., Хандеев В.И. Точный псевдополиномиальный алгоритм для одной задачи двухкластерного разбиения множества векторов // Дискрет. анализ и исслед. опер. — 2015. — Т. 22, № 4. — С. 50–62. RSCI (ядро РИНЦ). DOI: 10.17377/daio.2015.22.463.
Kel'manov A.V., Khandeev V.I. An Exact Pseudopolynomial Algorithm for a Problem of the Two-Cluster Partitioning of a Set of Vectors // J. Appl. Ind. Math. — 2015. — V. 9, № 4. — P. 497–502. Scopus. DOI: 10.1134/S1990478915040067.

[4] Кельманов А.В., Хандеев В.И. Полностью полиномиальная аппроксимационная схема для специального случая одной квадратичной евклидовой задачи 2-кластеризации // Журн. вычисл. математики и мат. физики. — 2016. — Т. 56, № 2. — С. 332–340. RSCI (ядро РИНЦ). DOI: 10.7868/S0044466916020113.
Kel'manov A.V., Khandeev V.I. Fully Polynomial-Time Approximation Scheme for a Special Case of a Quadratic Euclidean 2-Clustering Problem // Comput. Math. Math. Phys. 2016. — V. 56, № 2. — P. 334–341. Scopus, WoS. DOI: 10.1134/S0965542516020111.

[5] Кельманов А.В., Хамидуллин С.А., Хандеев В.И.. Точный псевдополиномиальный алгоритм для одной задачи разбиения последовательности // Автоматика и телемеханика. — 2017. — № 1. — С. 80–90. RSCI (ядро РИНЦ).
Kel'manov A.V., Khamidullin S.A., Khandeev V.I. Exact Pseudopolynomial Algorithm for one Sequence Partitioning Problem // Automat. Remote Control. — 2017. — V. 78, № 1. — P. 66–73. Scopus, WoS. DOI: 10.1134/S0005117917010052.

[6] Кельманов А.В., Хамидуллин С.А., Хандеев В.И. Полностью полиномиальная аппроксимационная схема для одной задачи двухкластерного разбиения последовательности // Дискрет. анализ и

исслед. опер. — 2016. — Т. 23, № 2. — С. 21–40. RSCI (ядро РИНЦ). DOI: 10.17377/daio.2016.23.511.
Kel'manov A.V., Khamidullin S.A., Khandeev V.I. A Fully Polynomial-Time Approximation Scheme for a Sequence 2-Cluster Partitioning Problem // J. Appl. Ind. Math. — 2016. — V. 10, № 2. — P. 209–219. Scopus. DOI: 10.1134/S199047891602006X.

[7] Кельманов А.В., Михайлова Л.В., Хамидуллин С.А., Хандеев В.И. Приближенный алгоритм для задачи разбиения последовательности на кластеры с ограничениями на их мощность // Тр. Ин-та математики и механики УрО РАН. — 2016. — Т. 22, № 3. — С. 144–152. RSCI (ядро РИНЦ). DOI: 10.21538/0134-4889-2016-22-3-144-152.

[8] Kel'manov A.V., Mikhailova L.V., Khamidullin S.A., Khandeev V.I. An Approximation Algorithm for a Problem of Partitioning a Sequence into Clusters with Restrictions on Their Cardinalities // Lect. Notes Comput. Sci. — 2016. — V. 9869. — P. 171–181. WoS, Scopus. DOI: 10.1007/978-3-319-44914-2_14.

[9] Кельманов А.В., Михайлова Л.В., Хамидуллин С.А., Хандеев В.И. Приближенный алгоритм для задачи разбиения последовательности на кластеры // Журн. вычисл. математики и мат. физики. — 2017. — Т. 57, № 8. — С. 149–157. RSCI (ядро РИНЦ). DOI: 10.7868/S0044466917080087.

Тезисы докладов

[10] Кельманов А.В., Хандеев В.И. 2-Приближенный полиномиальный алгоритм для одной задачи кластерного анализа // Материалы V Всерос. конф. «Проблемы оптимизации и экономические приложения». (Омск, 2–6 июля, 2012). — 2012. — С. 135.

[11] Кельманов А.В., Хандеев В.И. Полиномиальный алгоритм с оценкой точности 2 для решения одной задачи кластерного анализа // Интеллектуализация обработки информации: 9-я междунар. конф. Сб. докл. (Черногория, Будва, 16–22 сентября, 2012). — М.: Торус Пресс. — 2012. — С. 279–282.

[12] Кельманов А.В., Хандеев В.И. Рандомизированный алгоритм для одной задачи кластерного анализа // Материалы междунар. конф. «Дискретная оптимизация и исследование операций». (Новосибирск, Академгородок, 24–28 июня, 2013). — Новосибирск: ИМ СО РАН. — 2013. — С. 160.

[13] Кельманов А.В., Хандеев В.И. Рандомизированный алгоритм для одной NP-трудной задачи кластерного анализа // Математические методы распознавания образов: 16-я Всерос. конф. Тезисы докл. (Казань, 6–12 сентября, 2013). — М.: Торус Пресс. — 2013. — С. 35.

[14] Kel'manov A.V., Khandeev V.I. A randomized algorithm for a clustering problem // Proc. of IV Int. Conf. «Optimization and applications» (Petrovac, Montenegro, Sept. 22–28, 2013). — P. 86.

[15] Kel'manov A.V., Khandeev V.I. An exact pseudopolynomial algorithm for a two-cluster partitioning problem // Abstracts of the 16th Baikal Int. School-Seminar «Methods of Optimization and Their Applications». (Irkutsk, June 30 – July 6, 2014). — Irkutsk: Melentiev Energy Systems Institute SB RAS. — 2014. — P. 51.

[16] Kel'manov A.V., Khandeev V.I. An exact pseudopolynomial algorithm for a bi-partitioning problem // Abstracts of the V Int. Conf. «Optimization and Applications». (Petrovac, Montenegro, Sept. 28 – Oct. 4, 2014). — 2014. — P. 108–109.

[17] Kel'manov A.V., Khandeev V.I. An exact pseudopolynomial algorithm for a vectors set bi-partitioning problem // Abstracts of the 10th Int. Conf. «Intelligent Information Processing». (Greece, Crete, Oct. 4–11, 2014). — 2014. — P. 94–95.

[18] Кельманов А.В., Хамидуллин С.А., Хандеев В.И. Точный псевдополиномиальный алгоритм для одной задачи бикластеризации последовательности // XV Всерос. конф. «Математическое программирование и приложения». Тезисы докл. (Екатеринбург, 2–6 марта, 2015). — Екатеринбург: ИММ УрО РАН. — 2015. — С. 139–140.

[19] Кельманов А.В., Хандеев В.И. FPTAS для одной задачи двухкластерного разбиения множества векторов // XV Всерос. конф. «Математическое программирование и приложения». Тезисы докл. (Екатеринбург, 2–6 марта, 2015). — Екатеринбург: ИММ УрО РАН. — 2015. — С. 141–142.

[20] Kel'manov A.V., Khandeev V.I. FPTAS for special case of a quadratic Euclidean bi-partitioning problem // Abstract of the 28th Conf. of the European Chapter on Combinatorial Optimization. (Italy, Catania, May 28–30, 2015). — 2015. — P. 30.

[21] Кельманов А.В., Хамидуллин С.А., Хандеев В.И. FPTAS для специального случая одной квадратичной евклидовой задачи бикластеризации последовательности // Материалы VI междунар. конф. «Проблемы оптимизации и экономические приложения». (Омск, 28 июня – 4 июля, 2015). — 2015. — С. 138.

[22] Кельманов А.В., Хамидуллин С.А., Хандеев В.И. Полностью полиномиальная приближенная схема для одной задачи 2-кластеризации последовательности // Математические методы

распознавания образов: 17-я Всерос. конф. Тезисы докл. (Светлогорск, 19–25 сентября, 2015). — М.: Торус Пресс. — 2015. — С. 104–105.

[23] Кельманов А.В., Хандеев В.И. Полностью полиномиальная приближенная схема для одной квадратичной задачи 2-кластеризации // Математические методы распознавания образов: 17-я Всерос. конф. Тезисы докл. (Светлогорск, 19–25 сентября, 2015). — М.: Торус Пресс. — 2015. — С. 106–107.

[24] Kel'manov A.V., Khandeev V.I. Fully polynomial-time approximation scheme for a special case of a quadratic Euclidean 2-clustering problem // Abstracts of the VI Int. Conf. «Optimization and Applications». (Petrovac, Montenegro, Sept. 27 – Oct. 3, 2015). — 2015. — P. 94–95.

[25] Kel'manov A.V., Khamidullin S.A., Khandeev V.I. Fully polynomial-time approximation scheme for a sequence 2-clustering problem // Abstracts of the VI Int. Conf. «Optimization and Applications». (Petrovac, Montenegro, Sept. 27 – Oct. 3, 2015). — 2015. — P. 96–97.

[26] Kel'manov A.V., Khamidullin S.A., Khandeev V.I., Mikhailova L.V. An approximation algorithm for one NP-hard problem of partitioning a sequence into clusters with restrictions on their cardinalities // Abstracts of the VII Int. Conf. «Optimization and Applications». (Petrovac, Montenegro, Sept. 25 – Oct. 3, 2016). — 2016. — P. 80–81.

[27] Kel'manov A.V., Khamidullin S.A., Khandeev V.I., Mikhailova L.V. An approximation algorithm for a problem of partitioning a sequence into clusters // Abstracts of the VII Int. Conf. «Optimization and Applications». (Petrovac, Montenegro, Sept. 25 – Oct. 3, 2016). — 2016. — P. 82.

[28] Kel'manov A.V., Khamidullin S.A., Khandeev V.I., Mikhailova L.V. An approximation algorithm for one NP-hard problem of partitioning a sequence into clusters with restrictions on their cardinalities // Book of abstracts of the 11th Int. Conf. «Intelligent data Processing». (Barcelona, Spain, Oct. 10–14, 2016). — 2016. — P. 72–73.

[29] Kel'manov A.V., Khamidullin S.A., Khandeev V.I., Mikhailova L.V. An approximation algorithm for a problem of partitioning a sequence into clusters // Book of abstracts of the 11th Int. Conf. «Intelligent data Processing». (Barcelona, Spain, Oct. 10–14, 2016). — 2016. — P. 74–75.