Информационные процессы, Том 23, № 4, 2023, стр. 513–525.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ТЕХНИЧЕСКИХ СИСТЕМАХ

О задаче кластеризации с оптимизацией связей элементов кластеров

Марк Ш. Левин

Институт проблем передачи информации, Российская академия наук
Большой Каретный пер. 19, Москва 127051, Россия
email: mslevin@acm.org

Поступила в редколлегию 21.11.2023

Аннотация—В статье рассмотрены задачи кластеризации с комбинаторной точки зрения: (а) базовая задача кластеризации с оптимизацией связей элементов кластеров, (б) задача централизованной кластеризации с оптимизацией связей элементов кластеров, (в) задача кластеризации с оптимизацией связей элементов кластеров при наличии элементов нескольких типов, (г) связанные (близкие) задачи. Подход комбинаторной кластеризации является основой для материала статьи. Приводится обзор задач, методов решения и приложений. Рассмотрены оптимизационные модели для базовая задача кластеризации с оптимизацией связей элементов кластеров и задача кластеризации с многокритериальной оптимизацией связей элементов кластеров. Кратко описано приложение задаче кластеризации с оптимизацией связей элементов кластеров для задачи минимизации переключений обслуживания в мобильных беспроводных сетях. Числовые примеры иллюстрируют задачи и приложения.

Ключевые слова: кластеризация с ограничением на связи элементов кластеров, комбинаторная кластеризация, системы связи

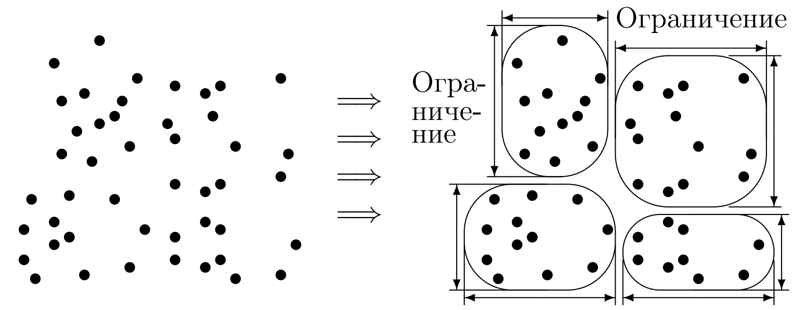
DOI: 10.53921/18195822_2023_23_4_513

1. ВВЕДЕНИЕ

Последние годы особое внимание уделяется различным комбинаторным задачам кластеризации [12,26,39,41,42,48,61,77,79]. Данная статья посвящена задаче кластеризации с оптимизацией связей элементов кластеров (capacitated clustering problem CCP) как специальному типу задач комбинаторной кластеризации [4, 15, 19, 32, 55, 58, 59, 63]. Эта задача известна также как специальный вид задачи разбиения графа [23, 62]. В данной работе задача рассматривается как компонент инженерного подхода к комбинаторной кластеризации [48, 49, 50, 51, 52].

Рассматриваемая задача кластеризации с оптимизацией связей элементов кластеров (CCP) заключается в формировании заданного числа кластеров (групп) элементов исходного множества так, чтобы суммы весов элементов каждого кластера удовлетворяли ограничениям и суммы полезности пар элементов внутри кластеров максимизировались. Данная задача принадлежит к классу NP-трудных задач комбинаторной оптимизации [4,63]. На Рис. 1 представлен иллюстративный пример для данной задачи (для 4-х кластеров).

Обычно исследуются следующие основные типы задачи кластеризации с оптимизацией связей элементов кластеров: (а) базовая постановка задачи [19,57,58,59,63]; (б) задача централизованной кластеризации с оптимизацией связей элементов кластеров (CCCP) [13,15,32,64,65,80]; (в) задача кластеризации с оптимизацией связей элементов кластеров при наличии элементов нескольких типов (MCCP) [4,68]; (г)гетерогенная задача кластеризации с оптимизацией связей элементов кластеров (HCCP) [66].



(а) исходные элементы (б) решение кластеризации

Рис. 1. Иллюстрация для задачи

В таблице 1 перечислены основные типы задач кластеризации с оптимизацией связей элементов кластеров, некоторые перспективные постановки и соответствующие литературные источники. Данная статья основана на предварительной публикации [53].

Таблица 1. Задачи кластеризации с оптимизацией связей элементов кластеров

Ном.	Задача	Источник
1.	Основные задачи:	
1.1.	Обзор о задачах и методах решения	[19]
1.2.	Базовая задача кластеризации с оптимизацией связей элементов кластеров	[55, 57, 58, 59, 63]
1.3.	Задача централизованной кластеризации с оптимизацией связей элементов кластеров	[13, 14, 15, 32, 60] [64, 65]
1.4.	Задача кластеризации с оптимизацией связей элементов кластеров при наличии элементов нескольких типов	[4, 68]
1.5.	Гетерогенная задача кластеризации с оптимизацией связей элементов кластеров	[66]
1.6.	Гетерогенная задача централизованной кластеризации с оптимизацией связей элементов кластеров	[66]
1.7.	Задача кластеризации большой размерности с оптимизацией связей элементов кластеров	[32]
1.8.	Задача кластеризации с оптимизацией связей элементов кластеров на дереве	[31]
2.	Перспективные постановки:	
2.1.	Многокритериальная ССР	[34], данная статья
2.2.	Многокритериальные постановки всех версий задач	
2.3.	Постановки всех версий задач в условиях неопределенности (стохастические модели, модели на основе размытых множеств)	

2. ОПИСАНИЕ ОСНОВНЫХ ЗАДАЧ

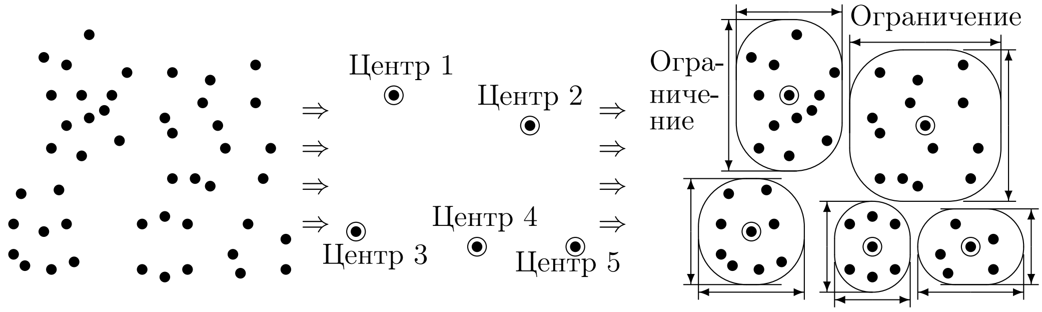
В базовой задаче кластеризации с оптимизацией связей элементов кластеров (ССР) рассматривается ненаправленный граф $G = (A, E)$ (A - множество вершин/узлов, E - множество дуг). Множество вершин разбивается на p различных (непересекающихся) групп (кластеров) [19, 55, 57, 58, 63, 83]. Таким образом решение кластеризации имеет вид $\widehat{X} = \{X_1, ..., X_k, ..., X_p\}$ (где X_k - кластер k) так что:

(1) сумма весов узлов в каждом кластере k ограничивается двумя ограничениями Q_k^- и Q_k^+ ($Q_k^- < Q_k^+$),

(2) суммы весов дуг внутри всех кластеров максимизируются.

Задача централизованной кластеризации с оптимизацией связей элементов кластеров (СССР) исследуется с работах [13, 15, 32, 60, 64, 65, 80]. Здесь рассматривается множество n точек в пространстве R^λ размерности $\lambda \geq 2$. Указанное множество точек разбивается на p различных групп (кластеров) с известной емкостью. Каждый кластер определяется цетройдом. В качестве целевой функции используется минимизация общего различия (total dissimilarity) в каждом кластере (т.е., сумма Евклидовых расстояний между точками и центром центройдом) при выполнении заданного ограничения на емкость кластера.

На Рис. 2 представлена иллюстрация для централизованной кластеризации с оптимизацией связей элементов кластеров (5 центройдов соответствуют кластерам). Центра для кластеров 1, 2, 3 выбраны на базе исходных элементов, центры для кластеров 4 и 5 определены дополнительно.



(а) исходные элементы (б) центры кластеров (в) решение кластеризации

Рис. 2. Иллюстрация для централизованной задачи кластеризации

Базовая СССР (или p-СССР) является нелинейной. Эта задача относится к классу NP-трудных комбинаторных задач. В последние годы исследуется специальный тип задачи: задача с p-центрами p-покупателями [10, 36, 70] (Таблица 2).

Таблица 2. Задачи с p-центрами и p-поставщиками при ограничениях

Ном.	Задача
1.	Задача p-поставщиков
2.	Задача p-центров, p-поставщиков и с ограничениями
3.	Задача с r-емкостью, p-поставшиками и p-центрами
4.	Балансная задача с p-поставщиками
5.	Хроматическая задача с p-поставщиками
6.	Задача p-поставщиков с условием живучести системы
7.	Задача с l-разнообразием и p-поставщиками

Балансная СССР p-центрами исследуется в [20].

Задача кластеризации с оптимизацией связей элементов кластеров при наличии элементов нескольких типов (multi-capacity clustering problem MCCP) была предложена в [4,68]. В данной задаче исследуются несколько типов элементов и для каждого типа элементов используется свой "ресурс"(емкость) и свое ограничение. В целом это усложняет задачу.

На Рис. 3 приведена упрощенная иллюстрация для MCCP (4 кластера и 4 типа элементов):

1. кластер 1: 6 элементов (4 элемента типа 1, 1 элемент типа 3, 1 элемент типа 4);

2. кластер 2: 7 элементов (5 элемента типа 2, 1 элемент 1, 1 элемент типа 4);

3. кластер 3: 8 элементов (1 элемент типа 1, 6 элементов типа 3, 1 элемент типа 4);

4. кластер 4: 8 элементов (1 элемент типа 1, 1 элемент типа 2, 6 элементов типа 4).

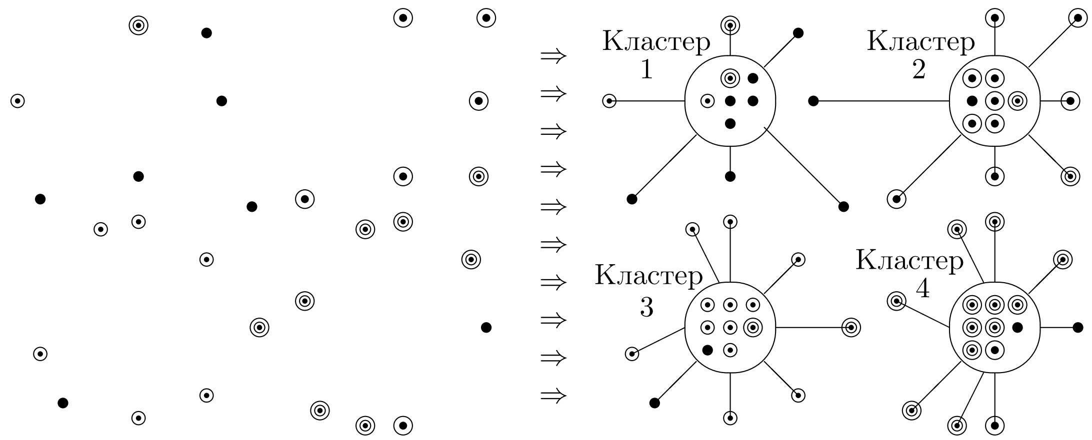


Рис. 3. Иллюстрация для MCCP

3. СВЯЗАННЫЕ ЗАДАЧИ

3.1. Основные связанный задачи

В литературе указаны следующие основные связанные (т.е., близкие) задачи для CCP:

1. CCP очень близка к задаче разбиения графа (graph partition problem GPP) [4, 83].

2. Близкими являются различные задачи кластеризации с ограничениями: (i) задачи кластеризации/разбиения графа с ограничениями [2, 5, 8, 9, 18, 29, 23, 27, 57, 62]; (ii) кластеризация с ограничением на размер кластеров [6, 38]; (iii) балансная кластеризация [3, 49, 54, 76].

3. CCP близка к задаче размещения оборудования с ограничениями [47].

4. CCP является эквивалентной задаче минимизации переключений обслуживания в мобильных беспроводных сетях (Handover Minimization Problem HMP) [58, 62].

5. Важной близкой задачей является задача максимизации разнообразий в группах (maximum diversity grouping problem MDGP) [46, 57, 72, 73, 75, 82, 83].

3.2. Задача максимизации разнообразий в группах

Задача максимизации разнообразий в группах (maximum diversity grouping problem MDGP) активно исследуется [22, 28, 46, 57, 69, 72, 73, 75, 82]:

Найти назначение множества элементов ($i \in A$) в различные (непересекающиеся) группы так, чтобы разнообразие между элементами в каждой группе (сумма парных расстояний $c_{i_1 i_2}$ $\forall i_1, i_2 \in A$ между всеми парами элементов в одной группе) было максимальным.

Здесь максимизируется разнообразие внутри групп. Очевидно, что MDGP представляет собой специальный случай CCP. MDGP называют задачей p-разбиения [22].

Эта задача входит в семейство задач разнообразия (family of diversity problems) [21, 28, 69].

3.3. Некоторые приложения

В Таблице 3 приведены приложения задачи кластеризации с оптимизацией связей элементов кластеров в промышленности и системах сервиса [25, 30, 56, 63, 65].

Таблица 3. Приложения

Ном.	Приложение	Источник
1.	Кластеризация сети с ограничением на ретрансляционные узлы для беспроводных сенсорных сетей с многими ретрансляционными узлами	[16]
2.	Разбиение узлов в рапределенных сетях связи	[4]
3.	Приложения в системах здравоохранения (например: распределение пациентов по больницам)	[4]
4.	Маршрутизация движущихся объектов	[24, 40, 83]
5.	Проектирование городских зон сбора отходов	[65]
6.	Планирование доставки почты	[7, 30, 63, 65]
7.	Размещение ИТ-бригад по организациям-разработчикам	[66]
8.	Проектирование интегральных схем (VLSI)	[56, 83]
9.	Формирование групп/бригад студентов	[4, 49, 78]
10.	Распределение продуктов в производственных системах переработки (очистки)	[4]

4. ОСНОВНЫЕ МЕТОДЫ РЕШЕНИЯ

Таблица 4 содержит список основные методов решения и соответствующие литературные источники.

5. ОПТИМИЗАЦИОННЫЕ МОДЕЛИ

5.1. Модель базовой задачи

Рассматриваются следующие компоненты задачи (CCP) [19, 55, 57, 58, 63, 83]:

(a) граф $G = (A, E)$ где A - множество n вершин (узлов, элементов) (т.е., $A = \{1, ..., i, ..., n\}$ и E - множество дуг $E = \{(i_1, i_2)\},\ \ \forall i_1, i_2 \in A$;

(б) вес вершины node $w_i \geq 0$ $(\forall i \in A)$;

(в) "полезность"(или вес) дуги c_{ij} $(\forall (i, j) \in E)$, следует отметить $c_{ij} = 0$ если дуга $(i, j)\overline{\in} E$;

(г) заданное число p различных (непересекающихся) кластеров $\{X_1 = (A_1, E_1), ..., X_k = (A_k, E_k), ..., X_p = (A_p, E_p)\}$ (т.е., $k = \overline{1,p}$, здесь $|A_{k_1} \& A_{k_2}| = 0 \quad \forall k_1, k_2 = \overline{1,k}$ и $k_1 \neq k_2$);

(е) два ограничения для каждого кластера $k = \overline{1,p}$: (i) минимальная емкость (т.е., ограничение) Q_k^- и (ii) максимальная емкость (т.е., ограничение) Q_k^+.

Математическая постановка для (CCP) обычно формулируется как модель квадратичного целочисленного программирования [11, 19, 83] с бинарными переменными: $x_{ik} = 1$ если элемент i помещен в кластер k и 0 в противном случае. Оптимизационная модель имеет вид:

$$\max \sum_{k=1}^{p} \sum_{i=1}^{n-1} \sum_{j>i}^{n} c_{ij}\ x_{ik}\ x_{jk} \tag{1.1}$$

$$s.t. \quad \sum_{k=1}^{p} x_{ik} = 1, \quad i = \overline{1,n}, \tag{1.2}$$

$$Q_k^- \leq \sum_{i=1}^{n} w_i x_{ik} \leq Q_k^+, \quad k = \overline{1,p}, \tag{1.3}$$

$$x_{ik} \in \{0, 1\}, \quad i = \overline{1,n}, \quad k = \overline{1,p} \tag{1.4}$$

Таблица 4. Основные методы решения

Ном.	Метод	Источник
1.	Обзоры, общие описания алгоритмов:	
1.1.	Обзор методов для CCP	[19]
1.2.	Библиографический обзор	[43]
1.3.	Обзор основных поисковых процедур	[83]
1.4.	Обзор эвристик и гибридных методов	[57, 58]
2.	Эвристики:	
2.1.	Метод с запретами (tabu search) (для CCP)	[83]
2.2.	Гибридный метод (метод и отжига метод с запретами) (для CCP)	[67]
2.3.	Эвристический поиск на основе разброса (scatter search heuristic) для CCP	[71]
2.4.	Метод пересоединения путей (path-relinking) с методом запретов (tabu search) (для CCCP)	[64]
2.5.	Интегрированный метод запретов (tabu search) и процедура случайного адаптивного поиска (Greedy Randomized Adaptive Search Procedure GRASP) для CCP	[57]
2.6.	Хаотичная (random) эвристика (для CCP)	[58]
3.	Специальные поисковые методы (например, локальный поиск, GRASP и др.):	
3.1.	Адаптивный генетический алгоритм с произвольным ключем и локальный поиск (для CCCP)	[15]
3.2.	Параллельный поиск (для CCP)	[60]
3.3.	Кластеризация с ограниченем на основе двойственного локального поиска	[33]
3.4.	Жадный (greedy) случайный (random) адаптивный поиск (для CCP)	[1]
3.5.	Реактивный GRASP с пересоединением путей (для CCP)	[19]
4.	Методы поиска с чередующимися окрестностями (Variable Neighborhood Search VNS):	
4.1.	VNS для CCP	[11, 45]
4.2.	Итеративный VNS для CCP	[44]
4.3.	Декомпозиционный VNS для CCP	[45]
4.4.	Итеративный VNS для CCCP	[80]
5.	Эволюционные методы:	
5.1.	Генетический алгоритм для CCP	[74]
5.2.	Гибридный эволюционный алгоритм для CCCP	[14]
5.3.	Метод дифференциальной эволюции для кластеризации с ограничениями	[35]
5.4.	Меметический алгоритм с запретами для CCP	[83]
5.5.	Мембранный эволюционный алгоритм для CCP	[56]
5.6.	Меметический алгоритм с привелегиями на основе декомпозиции для многокритериальной CCP	[34]
6.	Метаэвристики:	
6.1.	Мета-эвристическая схема (для HCCCP)	[66]
6.2.	Гибридная мета-эвристика (для CCCP)	[13]
6.3.	Гибридная метаэвристика (для MCCP)	[4]
6.4.	Метаэвристика (для CCP большой размерности - large-scale)	[32]
7.	Гибридные методы:	
7.1.	Гибридный алгоритм для CCP	[55]
7.2.	Гибридная эвристика для MCCP	[4]
7.3.	Гибридный эволюционный алгоритм для CCCP	[14]
7.4.	Гибридная метаэвристика для CCCP	[13]
7.5.	Трех-фазный гибридный алгоритм с использованием динамического развития (dynamic population) для MDGP	[82]
8.	Некоторые специальные методы:	
8.1.	Лагражева релаксация для многоразмерной версии CCP	[81]
8.2.	Математическая разрешимость с фиксированным параметром для CCP	[17]

Здесь целевая функция (1.1) соответствует общей "полезности"всех пар элементов, которые находятся в одном кластере. В данной модели ограничения играют следующие роли:

(i) ограничение (1.2) соответствует назначению каждого элемента в некоторый единственный кластер,

(ii) ограничение (1.3) соответствует следующему: сумма весов элементов в каждом кластере k ($k = \overline{1,p}$) находится между ограничениями: Q_k^- и Q_k^+ ($Q_k^- < Q_k^+$). Очевидно, что в упрощенном случае могут использоваться одинаковые ограничения для всех кластеров k: $Q_k^- = Q^-$ ($k = \overline{1,p}$) и $Q_k^+ = Q^+$ ($k = \overline{1,p}$).

5.2. Многокритериальная модель

Во-первых, в многоритериальной постановке вес дуги (т.е., "полезность") $c_{i,j}$ ($\forall (i,j) \in E$) трансформируется в вектор $\overline{c}_{i,j} = (c_{i,j}^1, ..., c_{i,j}^\gamma, ..., c_{i,j}^\mu)$. Получатся модель:

$$\max \sum_{k=1}^{p} \sum_{i=1}^{n-1} \sum_{j>i}^{n} c_{ij}^1 x_{ik}\ x_{jk}, ..., \max \sum_{k=1}^{p} \sum_{i=1}^{n-1} \sum_{j>i}^{n} c_{ij}^\gamma x_{ik}\ x_{jk}, ..., \max \sum_{k=1}^{p} \sum_{i=1}^{n-1} \sum_{j>i}^{n} c_{ij}^\mu x_{ik}\ x_{jk}; \quad (2.1)$$

$$s.t. \quad \sum_{k=1}^{p} x_{ik} = 1, \quad i = \overline{1,n} \quad (2.2)$$

$$Q_k^- \le \sum_{i=1}^{n} w_i x_{ik} \le Q_k^+, \quad k = \overline{1,p}, \quad (2.3)$$

$$x_{ik} \in \{0,1\}, \quad i = \overline{1,n}, \quad k = \overline{1,p} \quad (2.4)$$

Здесь следует проводить поиск Парето-эффективных решений.

Во-вторых, вес узла w_i ($i \in A$) может трансформироваться в вектор $\overline{w}_i = (w_i^1, ...w_i^\xi, ...w_i^\eta)$ и соответствующее ограничение для каждого кластера $k \in \{1, ..., p\}$ (т.е., ограничение (2.3)) меняется на набор η ограничений. Получается модель:

$$\max \sum_{k=1}^{p} \sum_{i=1}^{n-1} \sum_{j>i}^{n} c_{ij}^1 x_{ik}\ x_{jk}, ..., \max \sum_{k=1}^{p} \sum_{i=1}^{n-1} \sum_{j>i}^{n} c_{ij}^\gamma x_{ik}\ x_{jk}, ..., \max \sum_{k=1}^{p} \sum_{i=1}^{n-1} \sum_{j>i}^{n} c_{ij}^\mu x_{ik}\ x_{jk}; \quad (3.1)$$

$$s.t. \quad \sum_{k=1}^{p} x_{ik} = 1, \quad i = \overline{1,n} \quad (3.2)$$

$$Q_k^{\xi-} \le \sum_{i=1}^{n} w_i^\xi x_{ik} \le Q_k^{\xi+}, \ k = \overline{1,p}, \ \xi = \overline{1,\eta}, \quad (3.3)$$

$$x_{ik} \in \{0,1\}, \quad i = \overline{1,n}, \quad k = \overline{1,p} \quad (3.4)$$

Очевидно, что могут исследоваться и другие модификации многокритериальной модели.

6. О МИНИМИЗАЦИИ ПЕРЕКЛЮЧЕНИЙ ОБСЛУЖИВАНИЯ В МОБИЛЬНЫХ БЕСПРОВОДНЫХ СЕТЯХ

В последние годы рассматривается приложение задачи кластеризации с оптимизацией связей элементов кластеров к задач минимизации переключений обслуживания в мобильных беспроводных сетях [58, 62]. Иллюстрация для данного приложения представлена на Рис. 4: множество мобильных конечных пользователей, базовые станции и контроллеры для радио сети (RNC).

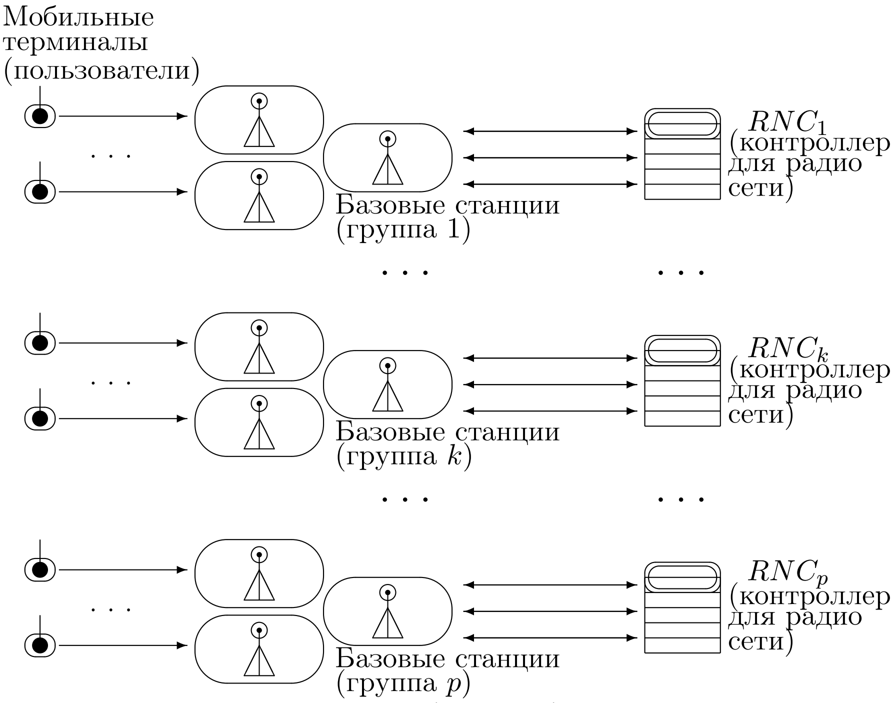


Рис. 4. Иллюстрация мобильной беспроводной сети

Контроллеры радио сетей осуществляют управление операциями базовых станций включая трафик и переключение обслуживания. Переключение между базовыми станцимяи, которые связаны с различными контроллерами (RNCs), приводит к нарушениям чаще, чем переключение между базовыми станциями связанными с одним контроллером. Задача минимизации переключений обслуживания в мобильных беспроводных сетях заключается в подключении (assignment) базовых станций к контроллерам. Множество базовых станций подключенных к одному контроллеру можно рассматривать как кластер. Задача минимизация числа переключении между различными кластерами является эквивалентной задаче максимизации числа переключений внутри кластеров.

Здесь используются следующие обозначения:

(i) множество базовых станций A = {1,...,i,...,n},

(ii) граф $G = (A, E)$, E - множество дуг (т.е., парных связей между базовыми станциями) G,

(iii) вес дуги c_u ($\forall u \in E$),

(iv) вес вершины/узла $w_i \geq 0$ ($\forall i \in A$),

(v) ограничения на емкость для каждого кластера Q^+ и Q^+.

Рассматриваемая задача эквивалентна задаче CCP:

Найти разбиение множества A (базовые станции) на p кластеров (т.е., групп базовых станций, каждая группа подключатся к одному контроллеру) таким образом, что:

(a) сумма весов дуг (c_u $\forall u \in E$) в кластере максимизируется,

(б) для каждого кластера сумма весов вершин ограничивается с двух сторон двумя ограничениями (w_i, $\forall i \in A$): Q^- и Q^+ ($Q^- < Q^+$).

7. ЗАКЛЮЧЕНИЕ

В статье рассмотрены версии задачи кластеризации с оптимизацией связей элементов кластеров: (а) базовая задача кластеризации с оптимизацией связей элементов кластеров, (б) задача централизованной кластеризации с оптимизацией связей элементов кластеров, (в) задача кластеризации с оптимизацией связей элементов кластеров при наличии элементов нескольких типов, (г) связанные (близкие) задачи, включая задачу максимизации разнообразий в группах. Приведен обзор задач, методов решения и приложений. Представлены оптимизационные модели для базовой задачи кластеризации с оптимизацией связей элементов кластеров и задачи кластеризации с многокритериальной оптимизацией связей элементов кластеров. Кратко описана задача минимизации переключений обслуживания в мобильных беспроводных сетях и показано ее сведение к рассматриваемой задачие кластеризации с оптимизацией связей элементов кластеров.

Представляются важными следующие перспективные направления: (1) исследование различных многокритериальных моделей для версий задачи кластеризации с оптимизацией связей элементов кластеров; (2) специальные рассмотрения версий задачи кластеризации с оптимизацией связей элементов кластеров в условиях неопределенности (например: стохастические постановки, постановки на основе применения размытых множеств); (3) построение новых стратегий решения (например, интегрированные гибридные методы); (4) дополнительные исследования возможных приложений задачи кластеризации с оптимизацией связей элементов кластеров; (5) использование задачи кластеризации с оптимизацией связей элементов кластеров в учебных курсах.

СПИСОК ЛИТЕРАТУРЫ

1. Ahmadi S., Osman I.H., Greedy random adaptive memory programming search for the capacitated clustering problems. Eur. J. Oper. Res., 2005, vol. 162, pp. 30–44.
2. Anand R., Reddy C.K., Graph-based clustering with constraints. In: Huang J.Z., Cao L., Srivastava J. (eds), Proc. of 15th Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining (PAKDD 2011), Part II, LNCS 6635, Springer, pp. 51–62, 2011.
3. Anaraki S.A.M., Haeri A., Soft and hard hybrid balanced clustering with innovative qualitative balancing approach. Inform. Sci., 2022, vol. 613, pp. 786–805.
4. de Araujo K.A.G., Guedes J.B., de Athayde Prata B., Hybrid metaheuristics for the multi-capacitated clustering problem. RAIRO-Oper. Res., 2022, vol. 56, pp. 1167–1185.
5. Banerjee A., Ghosh J., Scalable clustering algorithms with balancing constraints. Data Mining Knowl. Discov., 2006, vol. 13, pp. 365–395.
6. Baranwal M., Salapaka S.M., Clustering with capacity and size constraints: a deterministic approach. In: 2017 Indian Control Conf. (ICC), pp. 251–256, 2017.
7. Bard J.B., Jarrah A.I., Large-scale constrained clustering for rationalizing pickup and delivery operations. Transp. Res., Part B: Methodol., 2009, vol. 43, no. 5, pp. 542–561.
8. Basu S., Davidson I., Wagstaff K., Constrained Clustering: Advances in Algorithms, Theory and Applications. CRC Press, 2008.
9. Benlic U., Hao J.K., A multilevel memetic approach for improving graph k-partitions. IEEE Trans. Evolut. Comput., 2011, vol. 15, no. 5, pp. 624–642.
10. Bercea I.O., Gross M., Khuller S., Kumar A., Rosner C., Schmidt D.R., Schmidt M., On the cost of essentially fair clustering. In: Achlioptas D., Vegh L.A. (eds), Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques APPROX/Random 2019, pp. 18:1–18:22, 2019.
11. Brimberg J., Mladenovic N., Todosijevic R., Urosevic D., Solving the capacitated clustering problem with variable neighborhood search. Ann. Oper. Res., 2018, vol. 272, no. 1, pp. 289–321.

12. Cao Y., Wang Z., Combinatorial optimization-based clustering algorithm for wireless sensor networks. Math. Problems in Eng., 2020, art. ID 613704, pp. 1–13.

13. Chaves A.A., Lorena L., Clustering search algorithm for the capacitated centered clustering problem. Comp. Oper. Res., 2010, vol. 37, no. 3, pp. 552–558.

14. Chaves A.A., Lorena L., Hybrid evolutionary algorithm for the capacitated centered clustering problem. Exp. Syst. with Appl., 2011, vol. 38, pp. 5013–5018.

15. Chaves A.A., Goncalves J.F., Lorena L.A.N., Adaptive biased random-key genetic algorithm with local search for the capacitated centered clustering problem. Comp. Ind. Eng., 2018, vol. 124, pp. 331–346.

16. Chen D.-R., A link- and hop-constrained clustering for multi-hop wireless sensor networks. Comp. Commun., 2015, vol. 72, pp. 78–82.

17. Cohen-Addad V., Li J., On the fixed-parameter tractability of capacitated clustering. In: 42nd Conf. on Very Important Topics (CVIT 2016), art. 23, pp. 23:1–23:15, 2016.

18. Dang V.-T., Vu V.-V., Do H.-Q., Le T.K.O., Graph-based clustering with constraints and active learning. J. of Comp. Sci. Cybern., 2021, vol. 37, no. 1, pp. 71–89.

19. Deng Y., Bard J.F., A reactive GRASP with path relinking for capacitated clustering. J. of Heuristics, 2011, vol. 17, no. 2, pp. 119–152.

20. Ding H., Balanced k-center clustering when k is a constant. Electr. prepr., 6 p., Apr. 8, 2017. http://arxiv.org/abs/1704.02515 [cs.CG]

21. Duarte A., Sanchez-Oro J., Resende M., Glover F., Marti R., Greedy randomized adaptive search procedure with exterior path relinking for differential dispersion minimization. Inform. Sci., 2015, vol. 296, pp. 46–60.

22. Feo T., Goldschmidt O., Khellaf M., One-half approximation algorithms for the k-partition problem. Oper. Res., 1992, vol. 40, pp. S170–S173.

23. Ferreira C.E., Martin A., de Souza C.C., Weismantel R., Wolsey L.A., The node capacitated graph partitioning problem: a computational study. Math. Program., 1998, vol. 81, pp. 229–256.

24. Fisher M.L., Jaikumar R., A generalized assignment heuristic for vehicle routing. Networks, 1981, vol. 11, no. 2, pp. 109–124.

25. Franca P.M., Sosa N.M., Pureza V., An adaptive tabu search algorithm for the capacitated clustering problem. Int. Trans. Oper. Res., 1999, vol. 6, pp. 665–678.

26. Fuda N.J., Brejc K., Kruesi W.S., Ralston E.J., Bigley R., Shin A., Okada M., Meyer B.J., Combinatorial clustering of distinct DNA motifs directs synergistic binding of *Caenorhabditis elegans* dosage compensation complex to X chromosomes. PNAS, 2022, vol. 119, no. 37, art. e2211642119.

27. Galinier P., Boujbel Z., Fernandes M.C., An efficient memetic algorithm for the graph partitioning problem. Ann. Oper. Res., 2011, vol. 191, no. 1, pp. 1–22.

28. Gallego M., Duarte A., Laguna M., Marti R., Hybrid heuristics for the maximum diversity problem. Comput. Optim. Appl., 2009, vol. 44, no. 3, pp. 411–426.

29. Gancarski P., Dao T.-B.-H., Cremilleux B., Forestier G., Lampert T., Constrained clustering: current and new trends. In: Marquis P., Papini P., Prade H. (eds), A Guided Tour of Artificial Intelligence Research, Springer, pp. 447–484, 2020.

30. Geetha S., Poonthalir G., Vanathi P.T., Improved k-means algorithm for capacitated clustering problem. INFOCOMP J. Comput. Sci., 2009, vol. 8, pp. 52–59.

31. Gimadi E.Kh., Kurochkina A.A., Nagornaya E., On some effective algorithms for solving capacitated clustering and location problems on the tree and the real line. Applications. Irkutsk: ESI SB RAS, p. 95, 2017,

32. Gnagi M., Baumann P., A matheuristic for large-scale capacitated clustering. Comp. Oper. Res., 2021, vol. 132, art. 105304, pp. 1–15.

33. Gonzalez-Almagro G., Luengo J., Cano J.-R., Garcia S., DILS: constrained clustering through dual iterative local search. Comp. Oper. Res., 2020, vol. 121, art. 104979.

34. Gonzalez-Almagro G., Rosales-Perez A., Luengo J., Cano J.-R., Garcia S., ME-MEOA/D_{CC}: multiobjective constrained clustering through decomposition-based memetic elitism. Swarm Evolut. Comput., 2020, vol. 66, art. 100939.

35. Gonzalez-Almagro G., Luengo J., Cano J.-R., Garcia S., Enhancing instance-level constrained clustering through differential evolution. Appl. Soft Comput., 2021, vol. 108, art. 107435.

36. Goyal D., Jaiswal R., Tight FPT approximation for constrained k-center and k-supplier. Electr. prepr., 30 p., Oct. 27, 2021. http://arxiv.org/abs/2110.14242 [cs.DS]

37. Hansen P., Jaumard B., Cluster analysis and mathematical programming. Math. Progr., 1997, vol. 79, no. 1, pp. 191–221.

38. Hu C.W., Li H., Qutub A.A., Shrinkage clustering: a fast and size-constrained clustering algorithm for biomedical applications. BMC Bioinformatics, 2018, vol. 19, no. 1, art. 19, pp. 1–11.

39. Kim J., Lee W., Song J.J., Lee S.B., Optimized combinatorial clustering for stochastic processes. Cluster Comput., 2017, vol. 20, no. 2, pp. 1135–1148.

40. Koskosidis Y.A., Powell W.B., Clustering algorithms for consolidation of customer orders into vehicle shipments. Transp. Res., Part B: Methodol., 1992, vol. 26, no. 5, pp. 365–379.

41. Kumagai M., Komatsu K., Takano F., Araki T., Sato M., Kobayashi H., An external definition of the one-hot constraint and fast QUBO generation for high-performance combinatorial clustering. Int. J. Netw. Comput., 2021, vol. 11, no. 2, pp. 463–491.

42. Kumar V., Bass G., Tomlin C., Dulny J., Quantum annealing for combinatorial clustering. Quant. Inform. Proces., 2018, vol. 17, pp. 1–14.

43. Kuncheva L.I., Williams F.J., Hennessey S.L., A bibliographic view on constrained clustering. Electr. prepr., 18 p., Sep. 22, 2022. http://arxiv.org/abs/2209.11125 [cs.LG]

44. Lai X., Hao J.-K., Iterated variable neighborhood search for the capacitated clustering problem. Eng. Appl. of Artif. Intell., 2016, vol. 56, pp. 102–120.

45. Lai X., Hao J.-K., Fu Z.-H., Yue D., Neighborhood decomposition-driven variable neighborhood search for capacitated clustering. Comp. Oper. Res., 2021, vol. 134, art. 105362.

46. Lai X., Hao J.-K., Fu Z.-H., Yue D., Neighborhood decomposition based variable neighborhood search and tabu search for maximally diverse grouping. Eur. J. Oper. Res., 2021, vol. 289, no. 3, pp. 1067–1086.

47. Levin M.S., Modular System Design and Evaluation. Springer, 2015.

48. Levin M.S., On combinatorial clustering: literature review, methods, examples. J. of Commun. Technol. Electr., 2015, vol. 60, no. 12, pp. 1403–1428.

49. Levin M.S., On balanced clustering (indices, models, examples). J. of Commun. Technol. Electr., 2017, vol. 62, no. 12, pp. 1506–1515.

50. Levin M.S., Note on dominating sets problems. J. of Commun. Technol. Electr., 2021, vol. 66, no. Suppl. 1, pp. S8–S22.

51. Levin M.S., Clustering models based on graph edge coloring. J. of Commun. Technol. Electr., 2022, vol. 67, no. 12, 1570–1577.

52. Levin M.S., On the clique partitioning of a graph. J. of Commun. Technol. Electr., 2022, vol. 67, no. S2, pp. S267–S274.

53. Levin M.S., On capacitated clustering problem. Preprint, 14 p., Nov. 21, 2023. DOI: 10.13140/RG.2.2.33224.70401 (ResearchGate)

54. Lin Y., Tang H., Li Y., Fang C., Xu Z., Zhou Y., Zhou Z., Generating clusters of similar sizes by constrained balanced clustering. Appl. Intell., 2022, vol. 52, pp. 5273–5289.

55. Liu Y., Guo P., Zeng Y., HA-CCP: a hybrid algorithm for solving capacitated clustering problem. Comput. Intell. Neurosci., Volume 2022, art. 6400318, pp. 1–24, 2022

56. Liu Y., Guo P., Zeng Y., MEACCP: A membrane evolutionary algorithm for capacitated clustering problem. Inform. Sci., 2022, vol. 591, pp. 319–343.

57. Martinez-Gavara A., Campos V., Gallero M., Laguna M., Marti R., Tabu search and GRASP for the capacitated clustering problem. Computat. Optim. and Appl., 2015, vol. 62, no. 2, pp. 589–607.

58. Martinez-Gavara A., Landa-Silva D., Campos V., Marti R., Randomized heuristics for the capacitated clustering problem. Inform. Sci., 2017, vol. 417, pp. 154–168.

59. Mehrotra A., Trick M.A., Cliques and clustering: a combinatorial approach. Oper. Res. Lett., 1998, vol. 22, pp. 1–12.

60. Melo Morales D., Chaves A.A., Fazenda A.L., Parallel clustering search applied to capacitated centered clustering problem. In: 2019 IEEE Int. Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 542–548, 2019.

61. Mirkin B., Muchnik I., Combinatorial optimization in clustering. In: Du D.-Z., Pardalos P.M. (eds), Handbook of Combinatorial Optimization. vol. 2, Springer, pp. 261–329, 1999.

62. Moran-Mirabal L.F., Gonzalez-Velarde J.L., Resende M.G., Silva R.M., Randomized heuristics for handover minimization in mobility networks. J. of Heuristics, 2-13, vol. 19, no. 6, pp. 845–880.

63. Mulvey J.M., Beck M.P., Solving capacitated clustering problems. Eur. J. Oper. Res., 1984, vol. 18, pp. 339–348.

64. Muritiba A.E.F., Gomes M.J.N., de Souza M.F., Oria H.L.G., Path-relinking with tabu search for the capacitated centered clustering problem. Exp. Syst. with Appl., 2022, vol. 198, art. 116766.

65. Negreiros M., Palhano A., The capacitated centred clustering problem. Comput. and Oper. Res., 2006, vol. 33, pp. 1639–1663.

66. Negreiros M.J., Maculan N., Batista P.L., Rodrigues J.A., Palhano A.W.C., Capacitated clustering problems applied to the layout of the IT-teams in software factories. Ann. Oper. Res., vol. 316, no. 2, pp. 1157–1185.

67. Osman I.H., Christofides N., Capacitated clustering problems by hybrid simulated annealing and tabu search. Int. Trans. Oper. Res., 1994, vol. 1, pp. 317–336.

68. Prata B.A., The multi capacitated clustering problem. Technical report, Federal University of Ceara, Brazil, 2015.

69. Prokopyev O.A., Kong N., Martinez-Torres D.L., The equitable dispersion problem. Eur. J. Oper. Res., 2009, vol. 197, pp. 59–67.

70. Rosner C., Schmidt M., Privacy preserving clustering with constraints. In: Chatzigiannakis I., Kaklamanis C., Marx D., Sannella D. (eds), Int. Colloquium on Automata, Languages, and Programming ICALP 2018, pp. 96:1–96:14, 2018.

71. Scheuerer S., Wendolsky R., A scatter search heuristic for the capacitated clustering problem. Eur. J. Oper. Res., 2006, vol. 169, pp. 533–547.

72. Schulz A., A new mixed-integer programming formulation for the maximally diverse grouping problem with attribute values. Ann. Oper. Res., 2022, vol. 318, pp. 501–530.

73. Schulz A., The balanced maximally diverse grouping problem with integer attribute values. J. of Combin. Optim., 2023, vol. 45, art. 135, pp. 1–27.

74. Shieh H.-M., May M.-D., Solving the capacitated clustering problem with genetic algorithms. J. Chin. Inst. Ind. Eng., 2001, vol. 18, pp. 1–12.

75. Singh K., Sundar S., A new hybrid genetic algorithm for the maximally diverse grouping problem. Int. J. of Machine Learning and Cybernetics, 2019, vol. 10, no. 10, pp. 2901–2940.

76. Tang W., Yang Y., Zeng L., Zhan Y., Optimizing MSE for clustering with balanced size constraints. Symmetry, 2019, vol. 11, art. 338, pp. 1–16.

77. Vallet B., Soheilian B., Bredif M., Combinatorial clustering and its application to 3D polygonal traffic sigh reconstruction from multiple images. In: ISORS Annals of the Photogrammetry Remote Sensing and Spatial Information Sciences, vol. II-3, pp. 165–172, 2014.

78. Weitz R.R., Jelassi M.T., Assigning students to groups: a multi-criteria decision support system approach. Decis. Sci., 1992, vol. 23, no. 3, pp. 746–757.

79. Wierzchon S.T., Klopotek M.A., Algorithms of combinatorial cluster analysis. In: Modern Algorithms of Cluster Analysis, Studies in Big Data, vol. 34, Springer, pp. 67–161, 2018.

80. Xu Y., Guo P., Zeng Y., An iterative neighborhood local search algorithm for capacitated centered clustering problem. IEEE Access, 2022, vol. 10, pp. 34497–34510.

81. Yang Z., Chen H., Chu F., A lagrangian relaxation approach for a large scale new variant of capacitated clustering problem. Comput. Ind. Eng., 2011, vol. 61, pp. 430–435.

82. Yang X., Cai Z., Jin T., Tang Z., Gao S., A three-phase search approach with dynamic population size for solving the maximally diverse grouping problem Eur. J. Oper. Res., 2022, vol. 302, no. 3, pp. 925–953.

83. Zhou Q., Benlic U., Wu Q., Hao J.-K., Heuristic search to the capacitated clustering problem. Eur. J. Oper. Res., 2019, vol. 273, pp. 464–487.

On Capacitated Clustering Problem

Levin M.Sh.

The paper addresses capacitated clustering problems: (a) basic capacitated clustering problem, (b) capacitated centered clustering problem, (c) multi-capacity clustering problem, and (d) related problems. The paper material is based on combinatorial clustering viewpoint. A survey on the problems, solving approaches, and some applications is presented. The optimization models of basic capacitated clustering problem and multicriteria capacitated clustering problem are considered. An application of capacitated clustering as the handover minimization problem in mobile wireless networks is briefly described. Numerical examples illustrate problems and applications.

KEYWORDS: capacitated clustering, combinatorial clustering, communication systems