

Пономарев Андрей Васильевич

Модели и методы группировки объектов на основе  
взаимных отношений

05.13.01 — Системный анализ, управление и обработка информации  
(технические системы)

Научный руководитель: к.т.н., доцент Мустафин Н.А.

Санкт-Петербург — 2012

## Актуальность (I)

### Классические задачи и модели, связанные с группировкой объектов:

- кластерный анализ;
- некоторые задачи теории графов;
- задачи о покрытии и разбиении множества;
- дискретные задачи размещения.

### Известные обобщения:

- Amir - сведение к задаче разбиения графа;
- Faulkner - группирующий генетический алгоритм.

### Недостатки обобщений:

- рассмотрение только однородных объектов;
- высокая вычислительная сложность.

## Актуальность (II). Практический аспект

**Геолого-экономическое районирование** — выделение кластеров взаимосвязанных сырьевых, промышленных объектов и объектов инфраструктуры и определение стратегии недропользования через эти кластеры.

**Уровни геолого-экономического районирования:**

- промышленно-сырьевые узлы;
- геолого-экономические районы;
- геолого-промышленные комплексы.

**Сложности:**

- большое количество исходных объектов.

## Актуальность (II). Практический аспект

**Геолого-экономическое районирование** — выделение кластеров взаимосвязанных сырьевых, промышленных объектов и объектов инфраструктуры и определение стратегии недропользования через эти кластеры.

**Уровни геолого-экономического районирования:**

- промышленно-сырьевые узлы;
- геолого-экономические районы;
- геолого-промышленные комплексы.

**Сложности:**

- большое количество исходных объектов.

## Цель и задачи

**Цель** работы состоит в исследовании методов группировки объектов произвольной природы на основании отношений между этими объектами, с учетом структурных ограничений на состав групп и различных требований к покрытию образуемыми группами исходных множеств.

### Задачи

- создание методологического подхода к задаче группировки объектов, в особенности, централизованной группировки;
- разработка (или выбор) методов решения некоторых частных задач группировки;
- применение созданной платформы и методов для решения практической задачи формирования ПСУ.

## Классификация. Уточнение предмета

Подходы к группировке объектов в зависимости от способа целеполагания:

- процедурные;
- глобально оптимизирующие.

Виды группировки в зависимости от структуры исходных и результирующих множеств:

- однородная;
- неоднородная
  - централизованная;
  - составная.

## Классификация. Уточнение предмета

Подходы к группировке объектов в зависимости от способа целеполагания:

- процедурные;
- **глобально оптимизирующие.**

Виды группировки в зависимости от структуры исходных и результирующих множеств:

- однородная;
- неоднородная
  - **централизованная;**
  - составная.

## Формальная модель (I). Группа

Семейство исходных множеств  $\mathcal{O} = \langle O^{(1)}, O^{(2)}, \dots, O^{(k)} \rangle$ .

- Для централизованной группировки:  $\langle C, S \rangle$

**Базовая группа**  $G = \langle O_G^{(1)}, O_G^{(2)}, \dots, O_G^{(k)} \rangle$ ,  $O_G^{(i)} \subseteq O^{(i)}$ .

Операции: объединение ( $\cup$ ), пересечение ( $\cap$ ).

Отношения: ( $=$ ,  $\subset$ ,  $\subseteq$ ).

**Ограниченная группа (ОГ)** — базовая группа с ограничениями.

- Для централизованной группировки:  $|C(G)| = 1$ .

## Формальная модель (II). Группировка

**Группировка**  $\mathcal{G}$  — множество попарно не пересекающихся групп.

**Полнота** — характеристика покрытия группировкой  $i$ -того исходного множества.

- Для централизованной группировки: полная, полная по  $C$ , полная по  $S$ , неполная.

## Формальная модель (III). Система показателей

Оценка группировки  $val(\mathcal{G}) = f_{\mathcal{G}}(\{G\}, \xi)$

- $val(\mathcal{G}) = \sum_{G \in \mathcal{G}} val(G)$ .

Оценка группы  $val(G) = f_G(G, \zeta)$ .

- Простейшая СП:  $val(G) = \sum_{s \in S(G)} d(s, C(G))$

- СП со стоимостью активации:  
 $val(G) = \sum_{s \in S(G)} d(s, C(G)) - p(C(G))$

- СП с линейной стоимостью и бесплатным порогом:

$$val(G) = \sum_{s \in G(S)} d(s, C(G)) - a(C(G)) \cdot \max(0, -l(C(G)) + \sum_{s \in S(G)} r(s)).$$

## Формальная модель (IV). Резюме

Для определения задачи группировки необходимо задать:

- семейство исходных множеств;
- класс ограниченной группы — набор ограничений на соотношение объектов разных типов внутри группы;
- тип группировки по отношению к покрытию исходных множеств: полная, частично полная, неполная;
- систему показателей.

## Централизованная группировка в виде задачи ЦЛП

$$\text{Максимизировать } z = \sum_{i \in S} \sum_{j \in C} x_{ij} d(s_i, c_j) \quad (1)$$

при ограничениях

$$\sum_{j \in C} x_{ij} \stackrel{?}{=} 1 \quad i \in S \quad (2-a)$$

$$(y_j = 1) \quad j \in C \quad (2-б)$$

$$x_{ij} \leq y_j \quad i \in S, j \in C \quad (2-в)$$

$$x_{ij}, y_j \in \{0, 1\} \quad i \in S, j \in C$$

2-а, 2-б — ограничения покрытия (полнота по S и C);

2-в — ограничения целостности.

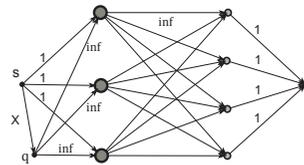
## Векторная природа объектов и дополнительные ограничения ОГ

Специфические виды ограничений:

- ограничение на разброс параметра в группе (содержание руды у месторождений одной группы не должно различаться более, чем на  $\Delta\%$ );
- количественные ограничения (в группу должно входить хотя бы одно месторождение с запасами более  $M$  тонн).

## Простейшая СП (I). Полная и полная по $C$ группировки

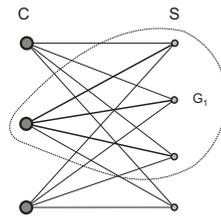
**Утверждение.** Полная и полная по  $C$  группировки с простейшей системой показателей сводятся к задаче нахождения потока минимальной стоимости в сети следующего вида:



$m \times n$  задача о назначении:

- Bertsekas D. Linear Network Optimization: Algorithms and Codes. The MIT Press, 1991

## Простейшая СП (II). Полная по $S$ и неполная группировки



**Утверждение.** Пара множеств  $(E, P)$  является матроидом, в качестве носителя в котором выступает  $E$  — множество ребер графа  $M$ , а независимым множеством является такое множество ребер графа  $M$ , ни одна пара из которых не имеет общей вершины в  $S$ .

Поиск оптимальной группировки  $\rightarrow$  поиск базы максимального веса в матроиде  $\rightarrow$  жадный алгоритм.

Сложность:  $O(|S||C| \cdot \log(|S||C|))$ .

## СП со стоимостью активации

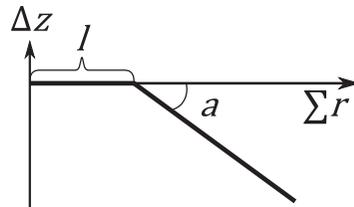
$$z = - \sum_{j=1}^m y_j p_j + \sum_{i=1}^n \sum_{j=1}^m x_{ij} d_{ij}. \quad (3)$$

Простейшая задача размещения (uncapacitated factory location problem).

Известные методы решения:

- методы неявного перебора (ветвей и границ)
- полиномиальные алгоритмы для частных случаев
- приближенные методы
  - LP-релаксация
  - поиск с запретами
  - генетические алгоритмы
  - метод роя частиц
  - и др.

## СП с линейной стоимостью и бесплатным порогом (I)



- $\Sigma r$  — суммарный ресурс группы;
- $\Delta z$  — влияние на целевую функцию;
- $l$  — порог;
- $a$  — коэффициент роста штрафа.

$$\text{Максимизировать } z = \sum_{j \in C} a_j q_j + \sum_{i \in S} \sum_{j \in C} x_{ij} d_{ij} \quad (4)$$

при ограничениях

$$q_j \leq 0$$

$$j \in C$$

$$q_j \leq y_j l_j - \sum_{i \in S} x_{ij} r_i$$

$$j \in C, i \in S$$

...

## СП с линейной стоимостью и бесплатным порогом (II)

### Гипотеза

Скорость решения задачи зависит от характеристик исходных данных.

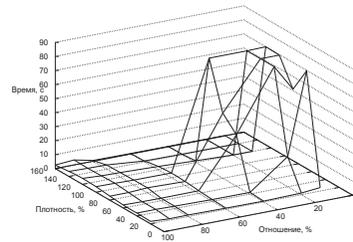
Выделенные параметры исходных наборов данных:

- *плотность* набора
- доля центров, у которых бесплатный порог больше 0
- *фаворитизм* — неравномерность распределения значений в отображении  $D$
- *r-d отношение*

## СП с линейной стоимостью и бесплатным порогом (III)

Этапы эксперимента с CPLEX:

1. Оценочный. 420 классов разных размеров
2. Уточняющий. 56 наборов по несколько экземпляров



**Итог:**

- быстро для большинства классов
- определенные сочетания плотности и доли центров с бесплатным порогом представляют сложность

## СП с линейной стоимостью и бесплатным порогом (IV)

Реализованные методы приближенного решения:

- локальный поиск
- локальный поиск со случайным стартом
- метод «лидер группы»
- вариант схемы GRASP

## Ускорение локального поиска

Параметры поиска:

- состояние:  $\{x\}_{n \times m}$ ;
- окрестность: REMOVE-ADD;
- стратегия: лучший сосед.

Структура данных «куча»:

- добавление/удаление:  $O(\log n)$ ;
- получение наибольшего:  $O(1)$ .

**Результат**

$$O(|C||S|) \rightarrow O(|S| \log |C|)$$

## Формирование ПСУ

Структура ПСУ:

- горно-обогатительный комбинат;
- набор месторождений одного геолого-промышленного типа.

Цепочка преобразований:

Добыча → транспорт → обогащение.

Способы учета действующего предприятия:

- не учитывается → простейшая СП;
- *псевдоприбыль* → СП со стоимостью активации;
- *без псевдоприбыли* → СП с линейной стоимостью и бесплатным порогом.

## Заклучение

В ходе работы получены следующие результаты:

- проведена формализация задачи группировки. Создан методический базис, с использованием которого можно классифицировать схожие задачи и выбирать наиболее подходящие методы решения;
- исследована двухуровневая централизованная группировка с различными системами показателей. Показана связь этих задач с известными задачами дискретной оптимизации, а для некоторых видов предложены оригинальные алгоритмы;
- разработаны программные модули для решения различных вариантов задачи группировки;
- задача формирования промышленно-сырьевых узлов представлена в виде задачи группировки и решена, с применением разработанных модулей, на примере данных о месторождениях Дальневосточного федерального округа.

Спасибо за внимание!  
Пожалуйста, вопросы.