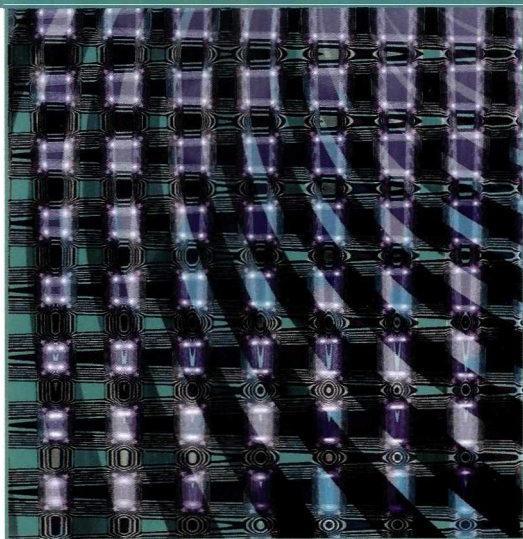




Л.А. Казаковцев

МЕТОД ЖАДНЫХ ЭВРИСТИК ДЛЯ ЗАДАЧ РАЗМЕЩЕНИЯ И КЛАСТЕРИЗАЦИИ



Министерство науки и высшего образования Российской Федерации
Сибирский федеральный университет

Л.А. КАЗАКОВЦЕВ

МЕТОД ЖАДНЫХ ЭВРИСТИК ДЛЯ ЗАДАЧ РАЗМЕЩЕНИЯ И КЛАСТЕРИЗАЦИИ

МОНОГРАФИЯ

Москва
ИНФРА-М
2024

УДК 519.6(075.4)
ББК 22.19
К14

*Работа выполнена при поддержке Министерства науки и высшего образования
Российской Федерации (грант № 075-15-2022-1121)*

Рецензенты:

Ступина А.А., доктор технических наук, заведующий кафедрой цифровых технологий управления Сибирского федерального университета;

Масич И.С., доктор технических наук, профессор кафедры системного анализа Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева

Казаковцев Л.А.

К14 Метод жадных эвристик для задач размещения и кластеризации :
монография / Л.А. Казаковцев. — Москва : ИНФРА-М, 2024. —
276 с. : ил.

ISBN 978-5-16-020386-7

В монографии излагается метод эффективной гибридизации алгоритмов локального поиска, стратегий глобального поиска, методов агрегирования данных, алгоритмов решения задачи Вебера с жадными агломеративными эвристическими процедурами в целях повышения точности и стабильности результатов решения дискретных и непрерывных задач кластерного анализа (автоматической группировки объектов) и задач размещения с различными мерами расстояния при большом объеме входных данных. Алгоритмы метода также эффективно применяются для решения некоторых задач монотонной псевдобулевой оптимизации. Кроме того, предложены алгоритмы решения задачи Вебера (задачи поиска центральной точки) с некоторыми специальными мерами расстояния.

Предназначена для научных работников, специалистов, студентов и аспирантов, занимающихся вопросами разработки алгоритмов кластерного анализа и алгоритмов для задач размещения.

553285

УДК 519.6(075.4)
ББК 22.19



ISBN 978-5-16-020386-7

© Казаковцев Л.А., 2024
© СФУ, 2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	6
1. СОВРЕМЕННОЕ СОСТОЯНИЕ РАЗВИТИЯ АЛГОРИТМОВ КЛАСТЕРНОГО АНАЛИЗА	10
1.1. Общая постановка задач автоматической группировки объектов и сферы их применения	10
1.2. Основные подходы к решению задач кластеризации	13
1.3. Задачи кластерного анализа и теория размещения	25
1.4. Стратегии глобального поиска в задачах размещения	32
1.5. Жадные эвристические процедуры в эволюционных алгоритмах ...	37
Результаты главы 1	40
2. ЖАДНЫЕ ЭВРИСТИЧЕСКИЕ АЛГОРИТМЫ В ЗАДАЧАХ МОНОТОННОЙ ПСЕВДОБУЛЕВОЙ ОПТИМИЗАЦИИ	42
2.1. Постановка задачи автоматической группировки на сети	42
2.2. Известные алгоритмы группировки узлов сети	45
2.3. Алгоритм метода изменяющихся вероятностей	48
2.4. Результаты метода изменяющихся вероятностей и подстройка параметров	55
2.5. Сравнительные результаты метода изменяющихся вероятностей	58
2.6. Результаты комбинированных методов	60
2.7. Параллельная версия алгоритма метода изменяющихся вероятностей	64
2.8. Жадная агрегативная эвристическая процедура для задач псевдобулевой оптимизации и некоторые свойства задач размещения	67
2.9. Постановка задачи составления расписаний загрузки производственных мощностей литейно-прокатного и химического производства	73
2.10. Алгоритм с жадной эвристикой для задачи о составлении расписания	79
2.11. Результаты вычислительных экспериментов для задачи о составлении расписания	87
Результаты главы 2	91

3. МЕТОД ЖАДНЫХ ЭВРИСТИК ДЛЯ НЕПРЕРЫВНЫХ И ДИСКРЕТНЫХ ЗАДАЧ АВТОМАТИЧЕСКОЙ ГРУППИРОВКИ ОБЪЕКТОВ.....	93
3.1. Общая постановка непрерывных задач автоматической группировки объектов	93
3.2. Известные методы для непрерывных задач размещения	99
3.3. Модификация жадной эвристической процедуры – частичное объединенное решение	105
3.4. Дальнейшие модификации жадной эвристики	110
3.5. Вычислительные эксперименты с модификациями жадных эвристик в составе генетического алгоритма	115
3.6. Применение новых модификаций жадных эвристических процедур к дискретным задачам размещения	121
3.7. Детерминированный алгоритм с жадной эвристической процедурой	130
3.8. Адаптивный метод жадных эвристик	135
3.9. Комбинация жадных эвристических алгоритмов с альтернативными алгоритмами локального поиска.....	142
3.10. Модификации для решения серии задач.....	146
3.11. Общая схема метода жадных эвристик.....	149
Результаты главы 3	156
4. ЗАДАЧИ АВТОМАТИЧЕСКОЙ ГРУППИРОВКИ И РАЗМЕЩЕНИЯ СО СПЕЦИАЛЬНЫМИ МЕРАМИ РАССТОЯНИЯ.....	158
4.1. Теория размещения и функции расстояний.....	159
4.2. Классы функций расстояния	163
4.3. Свойства задачи Вебера с прямоугольной метрикой.....	172
4.4. Алгоритм для задачи с метрикой лифта.....	178
4.5. Алгоритм для задачи с метрикой французской столицы (French Metro)	181
4.6. Применение и примеры для моделей размещения с метриками лифта и French metro, оценка вычислительной сложности алгоритмов	187
4.7. Обобщение метрик лифта и французской столицы, алгоритм для множественной задачи Вебера.....	191
4.8. Алгоритм для простейшей множественной задачи размещения.....	197
4.9. Задачи размещения с метриками, основанными на угловых расстояниях.....	202

4.10. Алгоритм для задачи Вебера с метрикой подъемного крана	206
4.11. Алгоритм для задачи Вебера с метрикой Москвы	211
4.12. Алгоритм для задачи размещения с метрикой британской железной дороги.....	216
4.13. Задача Вебера с мерой расстояния, включающей минимальную стоимость транспортировки.....	218
4.14. Задача Вебера на области, ограниченной дугами	230
Результаты главы 4.....	238
ЗАКЛЮЧЕНИЕ.....	239
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	242

ВВЕДЕНИЕ

Актуальность настоящей работы обусловлена ростом и бурным развитием систем машинного обучения, использующих, в частности, методы автоматической группировки данных, а также повышением общих требований к экономической эффективности, достигаемой, в том числе, за счет применения моделей оптимального размещения и автоматической группировки объектов по территориальному принципу или по сходству характеристик.

Прогресс в развитии методов кластерного анализа (автоматической группировки объектов и данных, наиболее популярные из которых основаны на решении р-медианной и подобных ей задач, таких как k-средних и k-медоид, в основном направлен на повышение их быстродействия. Поиск универсального и при этом точного метода решения подобных задач с большим объемом входных данных давно признан бесперспективным. Усилия исследователей сосредоточены на развитии компромиссных эвристических алгоритмов, дающих достаточно быстрое решение с минимальной ошибкой. Под эвристическим алгоритмом или процедурой, называемыми в специализированной литературе также словом «эвристика», будем понимать алгоритм (или его часть), не имеющий строгого обоснования, но дающий приемлемое решение задачи для большинства практических случаев. Наиболее распространены рандомизированные (использующие генератор случайных значений) процедуры, в связи с чем их результат зависит от значений случайных величин, от порядка следования данных и малых колебаний значений входных данных. В частности, рандомизированными алгоритмами, как правило, являются эволюционные (генетические) алгоритмы, основанные на применении эволюционных операторов – процедур, реализующих принципы биологической эволюции: селекции, скрещивания (рекомбинации), мутации. Эвристическими являются и так называемые «жадные» («пожирающие») алгоритмы, в которых на каждой итерации выбирается наилучшее решение из некоторого подмножества промежуточных решений. В то же время, некоторые из практических задач автоматической группировки требуют решения, которое не только было бы близко к точному решению задачи, но и являлось бы стабильным при многократных запусках алгоритма, воспроизводимым и, следовательно, проверяемым. Задачи должны решаться в интерактивном режиме при ограниченном времени работы с большим объемом входных данных. К таким задачам можно отнести, например, задачу формирования специальных партий электрорадиоизделий космического применения, где необходимость полу-

чения стабильных результатов обусловлена требованием воспроизводимости и проверяемости результатов расчетов, являющихся частью производственного процесса, в который вовлечены две стороны с не совпадающими интересами.

Наибольший вклад в теорию размещения и автоматической группировки объектов внесли Дрезнер, Бримберг, Младенович (задачи размещения), Хакими (задачи на сети), Весоловский (широкий круг задач), Лав (непрерывные задачи с различными метриками). Наиболее полные монографии по известным моделям и методам теории размещения составлены под редакцией Дрезнера и Хавахера (2004), Фарахани и Хекматфара (2009). Значительный вклад внесли Чен, Станимирович, Даскин, Питсулис, Пардалос, Резенде и др. Одними из первых в СССР модели размещения предприятий исследовали Черенин, Хачатуров. Модели размещения использовались и как основа моделей стандартизации и унификации, схожих с моделями, применяемыми в задачах автоматической группировки объектов, например, Институтом математики им. С.Л. Соболева СО РАН. Работы Береснева, Гимади, Колоколова а позже Кочетова, Забудского, Еремеева и др. создали фундамент для разработки программно-математического аппарата решения этих задач.

Детерминированные алгоритмы автоматической группировки и размещения существуют с середины XX века (работы Хакими). В то же время, вычислительная сложность этих алгоритмов, использующих, в частности, жадные агломеративные и диссоциативные эвристические процедуры, экспоненциально зависит от объема входных данных. Использование агломеративных жадных эвристических процедур в составе эволюционных алгоритмов для задач размещения позволяет получить хорошие по точности и стабильности результаты, но при этом эффективность алгоритма ограничена объемом входных, и их применение является обоснованным лишь для относительно узкого круга задач. Авторы исследований в этой области предлагают использование жадных эвристических процедур либо исключительно в виде самостоятельного алгоритма или самостоятельного эволюционного оператора (для дискретных задач размещения), либо в комбинации с отдельными известными методами локального поиска для данных задач. Тем не менее, точность и стабильность результатов получаемых жадными эвристическими процедурами, дает обоснованную надежду на построение эффективных алгоритмов, обладающих свойством стабильности получаемых результатов, пригодных к применению в интерактивных автоматизированных системах, предназначенных для решения задач автоматической группировки и размещения.

Для непрерывных задач размещения разработаны алгоритмы лишь для наиболее распространенных метрик и мер расстояния. В то же время, мера расстояния, используемая в практических задачах автоматической группировки объектов в географическом/геометрическом пространстве зависит от особенностей пространства и средств транспортировки. Этим обусловлена необходимость расширения арсенала используемых моделей и построения универсальных методов решения задач размещения с различными мерами расстояния (например, заданными алгоритмически), а также необходимость усовершенствования методов решения задач группировки с большим объемом входных данных.

В начатой работе излагается метод эффективной гибридизации алгоритмов локального поиска, стратегий глобального поиска, методов агрегирования данных, алгоритмов решения задачи Вебера с жадными агломеративными эвристическими процедурами в целях повышения точности и стабильности результатов решения дискретных и непрерывных задач автоматической группировки и размещения с различными мерами расстояния при большом объеме входных данных.

Результаты исследования дополняют арсенал эффективных эвристических методов решения NP-трудных задач автоматической группировки и размещения с широким кругом используемых мер расстояния. Разработаны новые виды жадных эвристических процедур и способы их комбинации со стратегиями глобального и локального поиска. Пополнен арсенал точных методов теории размещения, расширен круг используемых при этом моделей с нестандартными мерами расстояния, создана основа для синтеза новых моделей и методов. Кроме того, результаты систематизируют знания в области применения жадных агломеративных эвристических процедур в комбинации с другими известными методами оптимизации для решения задач автоматической группировки объектов, задач размещения и монотонной псевдобоулевой оптимизации.

Практическая ценность методов решения задач автоматической группировки и задач размещения обусловлена широтой их применения как непосредственно в практических задачах об оптимальном пространственном размещении и группировке объектов в географическом/геометрическом смысле (элементов городской инфраструктуры, точек доступа и базовых станций беспроводных сетей, коммутационного оборудования проводных сетей, складов, товаров на складах, точек сервиса, аварийных служб, датчиков мониторинга окружающей среды и др.), так и опосредованно в задачах автоматической группировки данных, кластерного анализа, теории оценивания и др. Разработанный метод позволяет по-

высить точность решения таких задач с большим объемом входных данных за ограниченное время, а также повысить стабильность получаемых результатов.

Для некоторых классов задач с нестандартными мерами расстояния разработан единый подход, основанный на декомпозиции в серию задач с прямоугольной метрикой. Дальнейшее обобщение задачи Вебера, а также решение ее особых случаев существенно расширяет возможности использования моделей и методов непрерывной теории размещения в практически важных задачах размещения и автоматической группировки.

Показано, что алгоритмы метода жадных эвристик применимы и в других областях, что продемонстрировано на практическом примере задачи оптимальной загрузки производственных мощностей литейно-прокатных производств или производств изделий из пластмасс с разнородной выпускаемой продукцией. Эффективность алгоритма при этом позволяет решать задачи в реальном времени, тогда как предложенные ранее алгоритмы не обеспечивают этой возможности.