



МОСКВА
"ФИНАНСЫ И СТАТИСТИКА"
1988

ББК 22.172
М23

Рецензенты: *Б. Г. Миркин, Ю. И. Тюрин*

Мандель И. Д.
М23 Кластерный анализ.— М.: Финансы и статистика.
1988.—176 с.: ил.
ISBN 5—279—00050—7

Освещается современное состояние кластерного анализа на основе сравнительного обсуждения многочисленных алгоритмов. Рассматривается методика применения кластерного анализа в социально-экономических исследованиях.

М $\frac{0702000000-103}{010(01)-88}$ 20—88

ББК 22.172

Научное издание

Мандель Игорь Давидович

КЛАСТЕРНЫЙ АНАЛИЗ

Зав. редакцией *Р. А. Казьмина*. Редактор *Л. В. Сергеева*
Мл. редакторы *В. Г. Крылова, Е. В. Гаврилова*
Худож. редактор *Ю. И. Артюхов*. Техн. редактор *Л. Г. Челышева*
Корректоры *Т. М. Колпакова, Т. Г. Кочеткова*
Обложка художника *С. Л. Витте*

ИБ № 2189

Сдано в набор 16.03.88. Подписано в печать 27.07.88. А11246. Формат 60×88¹/₁₆. Бум. кн.-журн.
Гарнитура «Литературная». Печать офсетная. Усл. п. л. 10,78. Усл. кр.-отт 11,04. Уч.-изд. л. 12,29.

Тираж 5 500 экз. Заказ 1115. Цена 1 р. 90 к.

Издательство «Финансы и статистика», 101000, Москва, ул. Чернышевского, 7.
Московская типография № 4 Союзполиграфпрома при Государственном комитете СССР по делам
издательств, полиграфии и книжной торговли. 129041, Москва, Б. Переяславская ул., д. 46.

ISBN 5—279—00050—7

© Издательство
«Финансы и статистика», 1988

Третье направление в точном решении задач классификации связано с недавно сформированной, но бурно развивающейся теорией монотонных систем [23, с. 29—34 и др.¹]. Упрощая понятие, дадим следующее определение: монотонной системой называется множество объектов, такое, что если из него изъять объект, то некоторая функция, определенная на любом подмножестве множества, не возрастет на любом подмножестве. Приведем удобный пример: на каждом объекте определена функция, ставящая ему в соответствие сумму расстояний от данного объекта до всех остальных; очевидно, что изъятие любого объекта может привести только к невозрастанию суммы расстояний для каждого объекта и для любого их подмножества.

Оказывается, в таких ситуациях можно строить точные и чрезвычайно быстрые алгоритмы оптимизации некоторым образом устроенных функционалов; их трудоемкость не превышает $O(N^2)$, тогда как для перечисленных выше методов точного решения она может достигать до экспоненциальной величины полного перебора. Правда, функционалы эти весьма специфичны. Так, для описанной системы с помощью построенного критерия можно выделять классы, расположенные как бы на концентрических сферах относительно центра множества; например, хорошо разделяются классы E и F на рис. 2.1. Более традиционные функционалы в духе табл. 2.5 для монотонных систем пока не разработаны. Можно, однако, использовать комбинированные алгоритмы — с помощью упомянутого функционала строить множество максимально удаленных точек — ядро, а затем, считая их центрами классов, проводить классификацию одним из эталонных алгоритмов. Такой способ первичного задания эталонов, как легко видеть, является обобщением эвристических процедур 45 и 46 из 2.2.

¹ Автор выражает признательность Е. П. Кузнецову и И. Б. Мучнику за обсуждение этой проблематики.

17. *Бородкин Л. И.* Многомерный статистический анализ в исторических исследованиях.—М.: МГУ, 1986.—188 с.
18. *Бухштабер В. М., Маслов В. И., Зеленюк Е. А.* Методы анализа и построения алгоритмов автоматической классификации на основе математических моделей//Прикладная статистика.—М.: Наука, 1983.— С. 126—144.
19. *Вапник В. Н.* Восстановление зависимостей по эмпирическим данным.—М.: Наука, 1979.—424 с.
20. *Васильев В. И.* Распознающие системы: Справочник.—Киев: Наукова думка, 1983.—424 с.

168

21. *Воронин Ю. А.* Теория классифицирования и ее приложения.—Новосибирск: Наука, 1985.—232 с.
22. *Воронин Ю. А., Гафуров Д. З., Шевченко Н. Г.* Некоторые классификационные вопросы распознавания.—Новосибирск: ВЦ СО АН СССР, 1983/Препринт № 380.—24 с.
23. *Всесоюзная конференция «Нечисловая статистика, экспертные оценки и смежные вопросы»:* Тезисы докладов.—М.-Таллин, 1984.—404 с.
24. *1 Всесоюзное совещание по статистическому и дискретному анализу нечисловой информации, экспертным оценкам и дискретной оптимизации:* Тезисы докладов.—М.-Алма-Ата: ВСНТО, 1981.—440 с.
25. *Гамбаров Г. М., Мандель И. Д., Рыбина И. А.* О некоторых метриках, возникающих в задачах обработки данных//Автоматика и телемеханика.—1980.— № 12.—С. 116—129.
26. *Гильбурд М. М.* Об эвристических методах решения задачи разбиения множества взаимосвязанных объектов//Автоматика и телемеханика.—1984.— № 1.— С. 107—113.
27. *Гришин В. Г.* Образный анализ экспериментальных данных.—М.: Наука, 1982.—240 с.
28. *Гуд И. Дж.* Ботриология ботриологии//Классификация и кластер.—М.: Мир, 1980.—С. 66—82.
29. *Диде Э. и др.* Методы анализа данных.—М.: Финансы и статистика, 1985.—360 с.
30. *Дорофеюк А. А.* Алгоритмы автоматической классификации: Обзор//Автоматика и телемеханика.—1971.— № 12.— С. 78—113.
31. *Дубровский С. А.* Прикладной многомерный статистический анализ.—М.: Финансы и статистика, 1982.—216 с.