

*Методы и модели анализа данных: OLAP и Data Mining/ Барсегян А. А.,
Куприянов М. С., Степаиенко В. В., Холод И. И. . — СПб.: БХВ-Петербург,
2004, - 336 с.: ил,*

ISBN 5-94157-522-X

В книге освещены основные направления в области анализа данных, организация хранилища данных, оперативный (OLAP) и интеллектуальный (Data Mining) анализ данных. Приведено описание методов и алгоритмов решения основных задач анализа: классификации, кластеризации и др. Описание идеи каждого метода дополняется конкретным примером его применения. Представлены стандарты и библиотека алгоритмов Data Mining. Прилагается компакт-диск, содержащий описание стандартов Data Mining, библиотеку алгоритмов Xelopes, лабораторный практикум и необходимое для практической работы программное обеспечение.

Содержание

Предисловие авторов.....	9
Data Mining и перегрузка информацией	11
Глава 1. Системы поддержки принятия решений	13
1.1. Задачи систем поддержки принятия решений.....	13
1.2. Базы данных — основа СППР.....	16
1.3. Неэффективность использования OLTP-систем для анализа данных	21
Выводы.....	26
Глава 2. Хранилище данных.....	27
2.1. Концепция хранилища данных.....	27
2.2. Организация ХД	34
2.3. Очистка данных	39
2.4. Хранилища данных и анализ	45
Выводы.....	45
Глава 3. OLAP-системы	49
3.1. Многомерная модель данных.....	49
3.2. Определение OLAP-систем	53
3.3. Концептуальное многомерное представление.....	54
3.3.1. Двенадцать правил Кодда	54
3.3.2. Дополнительные правила Кодда.....	55
3.3.3. Тест FASMI	57
3.4. Архитектура OLAP-систем.....	58
3.4.1. MOLAP	59
3.4.2. ROLAP	62
3.4.3. HOLAP	65
Выводы.....	66
Глава 4. Интеллектуальный анализ данных.....	67
4-1. Добыча данных — Data Mining	6
.....	6
7	
4.2. Задачи Data Mining	6
.....	6
8	
4 2 1 Классификация задач Data Mining	68
4 2 2 Задача классификации и регрессии	70

423	Задача поиска ассоциативных правил	72
424	Задача кластеризации	74
	Практическое применение Data Mining	76
43 i	Интернет-технологии	76
432	Торговля	76
433	Телекоммуникации	77
434	Промышленное производство	77
435	Медицина	78
436	Банковское дело	79
437	Страховой бизнес	80
438	Другие области применения	80
	Модели Data Mining	80
44 I	Предсказательные (predictive) модели	80
442	Описательные (descriptive) модели	83
	Методы Data Mining	83
4 5 1	Базовые методы	83
452	Нечеткая логика	83
453	Генетические алгоритмы	86
454	Нейронные сети	88
	Процесс обнаружения знаний	89
46 J	Основные этапы анализа	89
462	Подготовка исходных данных	91
	воды	91
	ша 5. Классификация регрессия	95
	Постановка задачи	95
	Представление результатов	96
52 I	Правила классификации	96
522	Деревья решений	97
523	Математические функции	99
	Методы построения правил классификации	99
5 3 1	Алгоритм построения 1-правил	99
5 3 2	Метод Naïve Bayes	101
	Методы построения деревьев решений	104
5 4 1	Методика "разделяй и властвуй"	104
5 4 2	Алгоритм покрытия	112

Методы построения математических функций	117
5 5 1 Общий вид	117
5 5 2 Линейные методы Метод наименьших квадратов	119
5 5 2 Нелинейные методы	120
5 5 3 Support Vector Machines (SVM)	121
Карта Кохонена	124
зоды	128
ша 6. Поиск ассоциативных правил	129
Постановка задачи	129
6 1 1 Формальная постановка задачи	129
6 1 2 Сиквенциальный анализ	Н2
6 1 3 Разновидности задачи поиска ассоциативных правил	135
6 2 Представление результатов	137
6 3 Алгоритмы	141
6 3 1 Алгоритм Арпол	141
6 3 2 Разновидности алгоритма Арпол	146
Выводы	147
Глава 7. Кластеризация	
.....	14
9	
7 1 Постановка задачи кластеризации	149
7 1 1 Формальная постановка задачи	152
7 1 2 Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации	154
7 2 Представление результатов	156
7 3 Базовые алгоритмы кластеризации	158
7 3 1 Классификация алгоритмов	158
7 3 2 Иерархические алгоритмы	159
Агломеративные алгоритмы	159
Дивизимные алгоритмы	161
7 3 3 Неиерархические алгоритмы	162
Алгоритм fc-means (Н«ird-c-means)	162
Алгоритм Fuzzy C-Means	166

Кластеризация по Гюстафсону-Кесселю	168
74 Кластеризация данных при помощи нечетких отношений	174
74] Анализ свойств нечетких бинарных отношений применительно	
к анализу данных	174
Отношения и свойства отношений	174
Сравнение данных	(79
Отношение α -толерантности	181
742 Отношение α -квазиэквивалентности	!82
Построение шкалы отношения α -квазиэквивалентности как алгоритм анализа данных	190
Об использовании шкалы α -квазиэквивалентности для анализа данных	191
Примеры анализа данных при помощи шкалы α -квазиэкви валентности	192
Выводы	204
Глава 8. Стандарты Data Mining 209	
81 Кратко о стандартах	209
8 2 Стандарт CWM	209
82 1 Назначение стандарта CWM	209
822 Структура и состав CWM	2i f
823 Пакет Data Mining	214
*3 Стандарт CRISP	218
S 3 1 Появление стандарта CRISP	218
8 3 2 Структура стандарта CRISP	218
8 3 3 Фазы и задачи стандарта CRISP	220
Стандарт PMML	225
Другие стандарты Data Mining	233
S 5 1 Стандарт SQL/MM	233
852 Стандарт OLE DB для Data Mining	235
853 СтандартJDMAPI	237
воды 237	
Глава 9. Библиотека Xelopes	241
Архитектура библиотеки	241

Диаграмма Model	244
9 2 1 Классы модели для Xelopes	244
922 Методы пакета Model	246
923 Преобразование моделей	247
Диаграмма Settings	248
9 3 1 Классы пакета Settings	248
9 3 2 Методы пакета Settings	250
Диаграмма Attribute	250
9 4 1 Классы пакета Attribute	250
942 Иерархические атрибуты	251
Диаграмма Algorithms	252
9 5 1 Общая концепция	252
952 Класс <i>MinmgAlgonthm</i>	253
953 Расширение класса <i>MinmgAlgonthm</i>	254
954 Дополнительные классы	256
955 Слушатели	256
Диаграмма DataAccess	256
9 6] Общая концепция	257
962 Класс <i>MmmgInputSiream</i>	258
963 Классы Mining векторов	258
964 Классы, расширяющие класс <i>MimngInputStream</i>	258
Диаграмма Transformation	259
Примеры использования библиотеки Xelopes	261
9 8 1 Общая концепция	261
982 Решение задачи поиска ассоциативных правил	264
983 Решение задачи кластеризации	266
984 Решение задачи классификации	268
воды 271	
Приложение 1. Нейронечеткие системы.....	273
1 Способы интеграции нечетких и нейронных систем	273
2 Нечеткие нейроны	277

3 Обучение методами спуска	279
4 Нечеткие схемы рассуждений	280
5 Настройка нечетких параметров управления с помощью нейронных сетей	286
6 Нейронечеткие классификаторы	293
Приложение 2. Особенности и эффективность генетических алгоритмов	
299	
П2 1 Методы оптимизации комбинаторных задач различной степени сложности	299
П2 2 Сущность и классификация эволюционных алгоритмов	304
П2 2 1 Базовый генетический алгоритм	304
П2 2 2 Последовательные модификации базового генетического алгоритма	305
П2 2 3 Параллельные модификации базового генетического алгоритма	307
f[2 3 Классификация генетических алгоритмов	310
П2 4 Особенности генетических алгоритмов, предпосылки для адаптации	311
П2 5 Классификация адаптивных ГА	314
П2 5 1 Основа адаптации	314
П2 5 2 Область адаптации	316
Адаптация на уровне популяции	316
Адаптация на уровне индивидов	317
Адаптация на уровне компонентов	318
П253 Основа управления адаптацией	318
П2 6 Двухнаправленная интеграция ГА и нечетких алгоритмов продукционного типа	319
Приложение 3. Описание прилагаемого компакт-диска	327
Список литературы	331
Предметный указатель	335