



Business intelligence
Modelli matematici e sistemi per le decisioni
Carlo Vercellis
McGraw-Hill, 2006

Indice

Introduzione	1
I Componenti del processo decisionale	7
1 Business intelligence	9
1.1 Decisioni efficaci e tempestive	9
1.2 Dati, informazioni e conoscenza	12
1.3 Ruolo dei modelli matematici	13
1.4 Architetture di business intelligence	14
1.4.1 Ciclo delle analisi di business intelligence	17
1.4.2 Fattori abilitanti per progetti di business intelligence	18
1.4.3 Realizzazione di un ambiente di business intelligence	19
1.5 Etica e business intelligence	22
1.6 Note bibliografiche	23
2 Sistemi di supporto alle decisioni	25
2.1 Definizione di sistema	25
2.2 Rappresentazione dei processi decisionali	27
2.2.1 Razionalità e problem solving	27
2.2.2 Il processo decisionale	30
2.2.3 Tipi di decisioni	32
2.2.4 Orientamenti nel processo decisionale	36
2.3 Evoluzione dei sistemi informativi	37
2.4 Definizione di sistema di supporto alle decisioni	39
2.5 Sviluppo di sistemi di supporto alle decisioni	42
2.6 Note bibliografiche	46
3 Data warehousing	47
3.1 Definizione di data warehouse	47
3.1.1 Data mart	50
3.1.2 Qualità dei dati	51
3.2 Architetture di data warehouse	52
3.2.1 Strumenti ETL	54
3.2.2 Metadati	55
3.3 Cubi e analisi multidimensionali	55
3.3.1 Gerarchie di concetti e operazioni OLAP	60
3.3.2 Calcolo dei cubi di dati	62
3.4 Note bibliografiche	63

II	Modelli e metodi matematici	65
4	Modelli matematici per le decisioni	67
4.1	Struttura dei modelli matematici	67
4.2	Fasi di sviluppo di un modello	69
4.3	Classi principali di modelli	72
4.4	Note bibliografiche	76
5	Data mining	77
5.1	Definizione di data mining	77
5.1.1	Modelli e metodi di data mining	79
5.1.2	Data mining, statistica classica e OLAP	80
5.1.3	Applicazioni di data mining	81
5.2	Rappresentazione dei dati in ingresso	82
5.3	Processo di data mining	83
5.4	Metodologie di analisi	88
5.5	Note bibliografiche	92
6	Preparazione dei dati	93
6.1	Validazione	93
6.1.1	Dati incompleti	94
6.1.2	Dati soggetti a rumore	95
6.2	Trasformazione	96
6.2.1	Standardizzazione	97
6.2.2	Estrazione di attributi	97
6.3	Riduzione	98
6.3.1	Campionamento	99
6.3.2	Selezione degli attributi	99
6.3.3	Analisi delle componenti principali	101
6.3.4	Discretizzazione	105
7	Esplorazione dei dati	109
7.1	Analisi univariata	109
7.1.1	Analisi grafica di attributi categorici	110
7.1.2	Analisi grafica di attributi numerici	112
7.1.3	Indici di posizionamento centrale per attributi numerici	114
7.1.4	Indici di dispersione per attributi numerici	116
7.1.5	Indici di posizionamento relativo per attributi numerici	121
7.1.6	Identificazione degli outlier per attributi numerici	123
7.1.7	Indici di eterogeneità per attributi categorici	125
7.1.8	Analisi della densità empirica	126
7.1.9	Informazioni di sintesi	129
7.2	Analisi bivariata	131
7.2.1	Analisi grafica	132
7.2.2	Indici di correlazione per attributi numerici	138
7.2.3	Tabelle di contingenza per attributi categorici	140
7.3	Analisi multivariata	143
7.3.1	Analisi grafica	143
7.3.2	Indici di correlazione per attributi numerici	145
7.4	Note bibliografiche	147

8	Regressione	149
8.1	Struttura dei modelli di stima	149
8.2	Regressione lineare semplice	152
8.2.1	Calcolo della retta di regressione	153
8.3	Regressione lineare multipla	156
8.3.1	Calcolo dei coefficienti di regressione	157
8.3.2	Assunzioni relative ai residui	159
8.3.3	Trattamento di attributi predittivi categorici	161
8.3.4	Regressione ridge	162
8.3.5	Regressione lineare generalizzata	163
8.4	Valutazione dei modelli di regressione	163
8.4.1	Normalità e indipendenza dei residui	164
8.4.2	Significatività dei coefficienti	165
8.4.3	Analisi della varianza	169
8.4.4	Coefficiente di determinazione	170
8.4.5	Coefficiente di correlazione lineare	171
8.4.6	Multi-collinearità delle variabili indipendenti	172
8.4.7	Limiti di confidenza e di predizione	172
8.5	Selezione delle variabili predittive	173
8.5.1	Esempio di sviluppo di un modello di regressione	176
8.6	Note bibliografiche	180
9	Serie storiche	181
9.1	Definizione di serie storica	181
9.1.1	Numeri indice	184
9.2	Valutazione dei modelli di serie storiche	185
9.2.1	Misure di distorsione	186
9.2.2	Misure di dispersione	186
9.2.3	Segnale di tracking	188
9.3	Analisi delle componenti	189
9.3.1	Media mobile	190
9.3.2	Scomposizione di una serie storica	191
9.4	Modelli di smoothing esponenziale	196
9.4.1	Smoothing esponenziale semplice	196
9.4.2	Smoothing esponenziale con correzione di tendenza	198
9.4.3	Smoothing con tendenza e stagionalità	200
9.4.4	Smoothing adattativo semplice	201
9.4.5	Smoothing a tendenza ridotta	202
9.4.6	Valori iniziali per i modelli di smoothing esponenziale	202
9.4.7	Eliminazione di tendenza e stagionalità	203
9.5	Modelli autoregressivi	204
9.5.1	Modelli a media mobile	205
9.5.2	Modelli autoregressivi a media mobile	205
9.5.3	Modelli autoregressivi integrati a media mobile	205
9.5.4	Identificazione di modelli autoregressivi	206
9.6	Combinazione di modelli predittivi	208
9.7	Scelta di un modello di previsione	209
9.7.1	Caratteristiche del processo previsionale	210
9.7.2	Selezione di una metodologia di previsione	211
9.8	Note bibliografiche	212

10	Classificazione	213
10.1	Problemi di classificazione	213
10.1.1	Tassonomia dei modelli di classificazione	216
10.2	Valutazione di modelli di classificazione	218
10.2.1	Metodo holdout	219
10.2.2	Campionamenti casuali ripetuti	220
10.2.3	Cross-validation	220
10.2.4	Matrici di confusione	221
10.2.5	Grafici ROC	224
10.2.6	Grafici di guadagno cumulato e di lift	225
10.3	Alberi di classificazione	228
10.3.1	Regole di separazione	231
10.3.2	Criteri di separazione univariati	234
10.3.3	Esempio di sviluppo di alberi di classificazione	238
10.3.4	Criteri di arresto e regole di pruning	242
10.4	Metodi bayesiani	243
10.4.1	Classificatore bayesiano naive	244
10.4.2	Esempio di utilizzo di un classificatore bayesiano naive	245
10.4.3	Reti bayesiane	247
10.5	Regressione logistica	248
10.6	Reti neurali	249
10.6.1	Perceptrone di Rosenblatt	250
10.6.2	Reti feed-forward a più livelli	251
10.7	Support vector machines	253
10.7.1	Minimizzazione del rischio strutturale	253
10.7.2	Iperpiani di margine massimo per la separazione lineare	257
10.7.3	Separazione nonlineare	261
10.8	Note bibliografiche	264
11	Regole associative	265
11.1	Struttura e valutazione delle regole associative	265
11.2	Regole associative a dimensione singola	269
11.3	Algoritmo Apriori	272
11.3.1	Generazione degli itemset frequenti	272
11.3.2	Generazione delle regole	273
11.4	Altre regole di associazione	275
11.5	Note bibliografiche	278
12	Clustering	279
12.1	Caratteristiche dei modelli di clustering	279
12.1.1	Tipologie di modelli di clustering	280
12.1.2	Misure di affinità	282
12.2	Metodi di partizione	288
12.2.1	Algoritmo delle K -medie	288
12.2.2	Algoritmo dei K -medoidi	291
12.3	Metodi gerarchici	292
12.3.1	Metodi gerarchici di agglomerazione	294
12.3.2	Metodi gerarchici di suddivisione	297
12.4	Valutazione dei modelli di clustering	298
12.5	Note bibliografiche	301

III Applicazioni di business intelligence	303
13 Modelli di marketing	305
13.1 Marketing relazionale	305
13.1.1 Motivazioni e obiettivi	307
13.1.2 Un ambiente di analisi di marketing relazionale	312
13.1.3 Lifetime value	315
13.1.4 Latenza dei modelli predittivi	317
13.1.5 Acquisizione	318
13.1.6 Retention	319
13.1.7 Cross-selling	320
13.1.8 Market basket analysis	321
13.1.9 Web mining	322
13.2 Gestione della forza di vendita	324
13.2.1 Processi decisionali nella gestione della forza di vendita	325
13.2.2 Modelli per la gestione della forza di vendita	327
13.2.3 Funzioni di risposta	328
13.2.4 Allocazione delle aree di vendita agli agenti	331
13.2.5 Pianificazione di visite e presentazioni	332
13.3 Esempi applicativi	336
13.3.1 Retention nelle telecomunicazioni	336
13.3.2 Acquisizione nei veicoli industriali	339
13.3.3 Cross-selling nel largo consumo	341
13.4 Note bibliografiche	344
14 Modelli logistici e produttivi	345
14.1 Ottimizzazione della supply chain	345
14.2 Modelli di ottimizzazione per la logistica	347
14.2.1 Pianificazione a medio termine	348
14.2.2 Capacità aggiuntiva	349
14.2.3 Risorse multiple	349
14.2.4 Modelli di backlog	350
14.2.5 Lotti minimi e costi fissi	352
14.2.6 Distinta base	353
14.2.7 Impianti multipli	354
14.3 Modelli di revenue management	355
14.3.1 Processi decisionali di revenue management	356
14.4 Esempi applicativi	358
14.4.1 Pianificazione nel settore alimentare	358
14.4.2 Pianificazione nel settore del packaging	365
14.5 Note bibliografiche	366
15 Data envelopment analysis	367
15.1 Misura di efficienza	367
15.2 Frontiera di efficienza	368
15.3 Il modello CCR	371
15.3.1 Definizione di obiettivi target	374
15.3.2 Riferimenti eccellenti	375
15.4 Individuazione di modi operativi efficienti	376
15.4.1 Analisi di cross-efficienza	376

15.4.2	Input e output virtuali	377
15.4.3	Restrizioni sui pesi	377
15.5	Altri modelli	377
15.6	Note bibliografiche	379
A	Strumenti software utilizzati	381
B	Dataset utilizzati	383
	Bibliografia	385
	Indice analitico	394