

# 目錄

	推薦序 .....	I
	序 .....	IV
<b>Chapter 1</b>	<b>資料採礦概念與應用</b>	
1.1	資料採礦的定義 .....	3
1.2	資料採礦的功能 .....	7
1.2.1	分類 .....	7
1.2.2	推估 .....	8
1.2.3	群集化 .....	8
1.2.4	同質分組 .....	9
1.2.5	序列 .....	11
1.2.6	描述 .....	11
1.3	資料採礦的應用 .....	13
<b>Chapter 2</b>	<b>資料採礦流程CRIPS-DM</b>	
2.1	CRISP-DM .....	20
2.2	定義商業問題 .....	23
2.3	定義分析資料 .....	25
2.2.1	時窗分析 .....	26
2.2.2	訓練組、鑑效組與測試組 .....	31
2.2.3	稀有事件 .....	34
2.4	資料預處理 .....	35
2.2.1	連續變數類別變數互換 .....	36
2.2.2	遺漏值 .....	37
2.2.3	極端值與標準化 .....	38
2.2.4	衍生變數與篩選有效變數 .....	40
<b>Chapter 3</b>	<b>SQL Server 2005資料採礦架構</b>	
3.1	SQL Server 2005 Analysis Services .....	44
3.1.1	Unified Dimension Model (UDM) .....	44
3.1.2	Analysis Services專案 .....	47
3.1.2	資料來源檢視 .....	52
3.1.3	資料視覺化 .....	61

3.2	SQL Server 2005 資料採礦架構 .....	66
3.3	對於微軟來說 什麼是「資料採礦」 .....	69

## Chapter 4 決策樹

4.1	決策樹演算法原理—分類樹 .....	73
4.1.1	決策樹演算法原理 .....	73
4.1.2	決策樹分岔準則計算 .....	75
4.1.3	決策樹修剪 .....	83
4.2	使用SQL 2005設計決策樹模型 .....	85
4.2.1	建立決策樹模型 .....	85
4.3	決策樹規則視覺化介面 .....	97
4.3.1	決策樹檢視器 .....	98
4.3.2	相依性網路 .....	101
4.3.3	採礦內容檢視器 .....	102
4.4	調整決策樹模型 .....	103
4.4.1	設定決策樹演算法參數 .....	103
4.4.2	複合決策樹 .....	107

## Chapter 5 群集演算法

5.1	群集演算法原理 .....	116
5.1.1	距離與標準化 .....	116
5.1.2	K-means演算法原理 .....	119
5.1.3	E-M演算法 .....	122
5.1.4	群集模型的資料預處理議題 .....	125
5.2	使用SQL Server 2005設計群集模型 .....	128
5.2.1	建立群集模型 .....	129
5.3	群集規則視覺化介面 .....	136
5.3.1	群集圖表 .....	136
5.3.2	群集設定檔 .....	138
5.3.3	群集特性 .....	141
5.3.4	群集辨識 .....	143
5.4	群集演算法調整 .....	144
5.4.1	調整群集演算法參數 .....	144
5.4.2	調整連續變數分配模式 .....	146

# 目錄

<b>Chapter 6</b>	<b>類神經網路</b>	
6.1	類神經網路原理 .....	153
6.1.1	人類神經元在做些什麼 .....	153
6.1.2	類神經元在做些什麼 .....	154
6.1.3	類神經網路架構與學習 .....	155
6.1.4	建構類神經網路注意事項 .....	161
6.2	使用SQL Server2005設計類神經網路模型 .....	162
6.3	類神經網路視覺化介面 .....	165
6.4	類神經網路模型調整 .....	170
<b>Chapter 7</b>	<b>關聯規則與時序群集</b>	
7.1	關聯規則原理 .....	176
7.1.1	關聯規則原理 .....	176
7.1.2	Apriori演算法 .....	179
7.1.3	時序群集原理 .....	182
7.2	使用SQL Server 2005設計關聯規則模型 .....	185
7.2.1	基本關聯規則 .....	185
7.2.2	利用關聯規則找出製程關鍵路徑 .....	191
7.2.3	建置時序群集規則 .....	196
7.3	關聯模型視覺化介面 .....	201
7.3.1	項目集 .....	201
7.3.2	規則 .....	202
7.3.3	相依性網路 .....	204
7.3.4	製程良率分析規則解讀 .....	206
7.4	時序群集模型視覺化介面 .....	209
7.4.1	群集圖表 .....	209
7.4.2	群集設定檔 .....	211
7.4.3	群集特性 .....	212
7.4.4	群集辨識 .....	213
7.4.5	狀態轉換 .....	214
7.5	關聯規則與時序群及模型調整 .....	216
7.5.1	關聯規則參數 .....	216
7.5.2	時序群集演算法參數 .....	217

<b>Chapter 8</b>	<b>時間序列</b>	
	8.1 時間序列原理 .....	221
	8.1.1 時間序列原理 .....	221
	8.2 使用SQL Server 2005設計時間序列模型 .....	226
	8.2.1 基本時間序列 .....	226
	8.2.2 時間序列交叉預測 .....	230
	8.3 時間序列規則視覺化介面 .....	233
	8.3.1 決策樹 .....	233
	8.3.2 時間序列圖表 .....	235
	8.4 調整時間序列模型 .....	238
	8.4.1 時間序列演算法參數 .....	238
	8.4.2 顯示歷程紀錄預測 .....	240
<b>Chapter 9</b>	<b>貝氏機率分類</b>	
	9.1 貝氏機率分類原理 .....	245
	9.1.1 貝氏機率分類原理 .....	245
	9.2 使用SQL Server 2005設計貝氏機率分類模型 .....	250
	9.2.1 建置貝氏機率分類 .....	250
	9.3 貝氏機率分類視覺化介面 .....	257
	9.3.1 相依性網路 .....	257
	9.3.2 屬性設定檔 .....	258
	9.3.3 屬性特性 .....	259
	9.3.4 屬性辨識 .....	260
	9.4 調整貝氏機率分類模型 .....	262
	9.4.1 貝氏機率分類演算法參數 .....	262
<b>Chapter 10</b>	<b>線性迴歸、迴歸樹與羅吉斯迴歸</b>	
	10.1 迴歸原理 .....	264
	10.1.1 迴歸基礎觀念 .....	264
	10.1.2 多元線性迴歸 .....	271
	10.1.3 迴歸樹 .....	274
	10.1.4 羅吉斯迴歸 .....	275
	10.2 使用SQL Server 2005設計迴歸模型 .....	278
	10.2.1 線性迴歸 .....	280
	10.2.2 迴歸樹 .....	286

# 目錄

	10.2.3 羅吉斯迴歸 .....	286
10.3	採礦模型視覺化介面 .....	287
	10.3.1 決策樹檢視器 .....	287
	10.3.2 相依性網路檢視器 .....	291
	10.3.3 羅吉斯迴歸類神經網路檢視器 .....	292
	10.3.4 模型內容檢視器 .....	294
10.4	迴歸演算法調整 .....	296
	10.4.1 迴歸演算法參數 .....	296
	10.4.2 羅吉斯迴歸演算法參數 .....	297
<b>Chapter 11</b>	<b>文字資料採礦</b>	
11.1	SSIS文字資料採礦預處理功能 .....	302
	11.1.1 詞彙擷取 .....	302
	11.1.2 詞彙查閱 .....	306
11.2	解析文字 .....	307
	11.2.1 文件詞彙擷取範例 .....	307
	11.2.2 文件詞彙查閱範例 .....	313
11.3	建立文字資料採礦模型 .....	318
	11.3.1 文件分群 .....	319
	11.3.2 文件分類 .....	322
	11.3.3 關鍵字關聯規則 .....	325
<b>Chapter 12</b>	<b>OLAP資料採礦</b>	
12.1	整合的多維度分析採礦解決方案 .....	330
	12.1.1 OLAP與資料採礦資料結構 .....	330
	12.1.2 Analysis Services 2005 OLAP採礦演算法 .....	332
12.2	使用OLAP設計分類模型 .....	333
	12.2.1 建置OLAP決策樹模型 .....	333
	12.2.2 採礦維度 .....	342
12.3	使用OLAP設計關聯規則模型 .....	348
	12.3.1 建置關聯規則規則 .....	348
<b>Chapter 13</b>	<b>資料採礦模型評估</b>	
13.1	資料採礦模型評估的錯誤觀念 .....	356
13.2	分類問題驗證－分類矩陣 .....	358

	13.3 增益圖 .....	365
	13.4 收益圖 .....	370
	13.5 散佈圖 .....	375
	13.6 其它模型評估 .....	376
	13.7 難以量化的資料採礦效益.....	376
<b>Chapter 14</b>	<b>與SSIS整合資料採礦流程</b>	
	14.1 SQL Server 2005資料整合服務.....	378
	14.1.1 SSIS物件 .....	380
	14.1.2 控制流程.....	381
	14.1.3 資料流程.....	384
	14.2 運用SSIS設計資料預處理流程 .....	387
	14.2.1 資料抽樣.....	387
	14.2.2 產生衍生變數.....	395
	14.3 運用SSIS設計資料採礦應用流程 .....	403
	14.3.1 採礦模型預測.....	403
	14.3.2 採礦模型處理.....	408
	14.3.3 SSIS自動化.....	409
<b>Chapter 15</b>	<b>DMX預測查詢</b>	
	15.1 預測查詢基本結構.....	417
	15.1.1 SELECT陳述式.....	419
	15.1.2 來源資料查詢.....	419
	15.1.3 預測連結.....	420
	15.2 基本預測查詢 .....	425
	15.2.1 Prediction Query .....	425
	15.2.2 Natural Prediction Join .....	431
	15.3 其他演算法預測查詢 .....	433
	15.3.1 群集演算法查詢 .....	433
	15.3.2 關聯規則演算法查詢.....	435
	15.3.3 時序群集演算法查詢.....	445
	15.3.4 時間序列演算法查詢.....	448
	15.4 預測函數 .....	452
	15.4.1 統計函數.....	452
	15.4.2 巢狀表格函數.....	455

# 目錄

	15.4.3 篩選函數.....	461
	15.4.5 其它函數.....	463
15.5	進階預測查詢 .....	465
	15.5.1 子查詢 .....	465
	15.5.2 排序與篩選 .....	467
	15.5.3 進階資料來源.....	468
	15.5.4 呼叫VBA與EXCEL函數 .....	472
<hr/>		
Chapter 16	結構定義查詢與OLE DB for Data Mining	
<hr/>		
16.1	OLE DB for Data Mining.....	480
	16.1.1 Mining Service Schema Rowset.....	481
	16.1.2 Mining Service Parameter Schema Rowset .....	488
	16.1.3 Mining Function Schema Rowset.....	490
	16.1.4 Mining Structure Schema Rowset .....	492
	16.1.5 Mining Structure Column Schema Rowset.....	493
	16.1.6 Mining Model Schema Rowset.....	499
	16.1.7 Mining Column Schema Rowset .....	501
	16.1.8 Mining Model Content Schema Rowset.....	506
16.2	模型內容查詢 .....	510
	16.2.1 模型內容查詢.....	511
	16.2.2 Reporting Services決策樹展現 .....	516
	16.2.3 模型欄位屬性查詢 .....	523
	16.2.4 模型鑽研.....	524
16.3	資料採礦系統預存程序 .....	525
	16.3.1 公用資料採礦預存程序 .....	526
	16.3.2 貝氏機率分類專用預存程序.....	527
	16.3.3 類神經網路專用預存程序 .....	529
	16.3.4 群集專用預存程序 .....	532
	16.3.5 相依性網路專用預存程序 .....	534
<hr/>		
Chapter 17	DMX資料定義語言與操作語言	
<hr/>		
17.1	資料採礦資料定義語言CREAT、ALTER、DROP .....	537
	17.1.1 建立採礦結構.....	538
	17.1.2 在採礦結構加入採礦模型 .....	542

	17.1.3 直接建立採礦結構與模型 .....	545
	17.1.3 複製模型定義 .....	547
	17.1.4 巢狀模型定義 .....	548
	17.1.5 刪除採礦模型 .....	549
	17.1.6 匯出匯入採礦模型定義 .....	549
17.2	資料採礦資料操作語言 .....	551
	17.2.1 訓練採礦模型 .....	551
	17.2.2 巢狀模型訓練 .....	553
	17.2.3 清空採礦模型 .....	553
	17.2.4 模型內容變更 .....	556
<hr/>		
<b>Chapter 18</b>	<b>資料採礦應用開發</b>	
<hr/>		
18.1	資料採礦預存程序－ADOMDSERVER .....	557
	18.1.1 計算關聯規則數量 .....	558
	18.1.2 搜尋最小值所在的迴歸樹節點 .....	562
	18.1.3 計算相關係數與共變異數 .....	568
	18.1.3 自訂DMX函數 .....	570
	18.1.4 產生DMX Create語法 .....	572
18.2	ADOMD Client .....	577
18.3	AMO .....	581
	18.3.1 AMO基礎操作 .....	583
	18.3.2 利用AMO追蹤模型處理進度 .....	584
<hr/>		
<b>Chapter 19</b>	<b>PMML預測模型標示語言</b>	
<hr/>		
19.1	存取PMML文件 .....	593
19.2	應用PMML文件 .....	595
19.3	PMML文件規格 .....	598
	19.3.1 標頭 .....	599
	19.3.2 Data Dictionary .....	600
	19.3.3 演算法標記 .....	602
	19.3.4 Mining Schema .....	604
	19.3.5 統計量 .....	605
	19.3.6 模型內容－model .....	608



# 目錄

<b>Chapter 20</b>	<b>Excel資料表分析工具</b>	
20.1	安裝資料採礦增益集 .....	611
20.2	Excel資料表分析工具 .....	617
20.1.1	分析關鍵因數 .....	620
20.1.2	偵測類別目錄 .....	625
20.1.3	根據範例填滿 .....	631
20.1.4	預測 .....	634
20.1.5	反白顯示例外狀況 .....	638
20.1.6	狀況分析-搜尋目標 .....	641
20.1.7	狀況分析-假設 .....	643
<b>Chapter 21</b>	<b>Excel資料採礦用戶端</b>	
21.1	資料準備 .....	648
21.1.1	瀏覽資料 .....	649
21.1.2	清除資料 .....	653
21.1.3	分割資料 .....	661
21.2	資料模型化 .....	667
21.2.1	資料模型化 .....	667
21.2.2	進階建模 .....	671
21.2.3	將模型加入結構 .....	674
21.3	精確度與驗證 .....	676
21.3.1	採礦精確度圖表 .....	676
21.3.2	分類矩陣 .....	679
21.3.3	收益圖 .....	680
21.4	模型使用方式與管理 .....	681
21.4.1	模型瀏覽 .....	681
21.4.2	模型查詢 .....	682
21.4.3	管理模型 .....	685
21.4.4	追蹤 .....	687
<b>Chapter 22</b>	<b>Visio資料採礦範本</b>	
22.1	決策樹範本 .....	692
22.1.1	決策樹範本 .....	692
22.2	群集範本 .....	704
22.3	相依性網路範本 .....	712