

目次

第1章 はじめに	1
第2章 ビジネス課題	9
イントロダクション	9
プロジェクトを実行可能なものとするための指針	10
データの利用可能性におけるプロジェクトの実行可能性 ——特殊な検討事項	11
プロジェクトの利益に影響する因子	12
プロジェクトのコストに影響を与える因子	13
例1：カスタマーコールセンター——目的：顧客クレームへの対応	14
ストロング氏のプロジェクトにおける利益およびコストの総括	16
例2：オンラインミュージックアプリの事例 ——目的：モバイルデバイスにおける広告効果	17
メロディオンライン社のプロジェクトに関する利益とコストの総括	19
まとめ	21
もっと知りたい読者のために	21
第3章 さまざまなデータソースや情報を組み合わせる	23
イントロダクション	23
製品やサービスに関するデータ	25

サーベイとアンケート	27
サーベイとアンケート：データテーブルの実装について	32
フォームを設計する際に気をつけること	34
ポイントカード/お客様カード	35
ポイントカードの登録フォーム：データテーブルの構成	40
デモグラフィックデータ	48
国勢調査（2010年のアメリカ国勢調査データより）	49
マクロ経済データ	51
競合についてのデータ	55
株式、シェア、コモディティ、投資などの金融マーケットデータ	57
第4章 データ表現	61
イントロダクション	61
基本的なデータ表現	61
基本的なデータ型	61
異なる型の変数の表現，比較，処理	64
変数の主な型	65
変数に含まれる値の標準化	70
変数に格納された値の分布	71
異常値（外れ値）	72
発展的なデータ表現	75
階層型データ	76
セマンティックネットワーク	77
グラフデータ	78
ファジーデータ	80
第5章 データの質	83
イントロダクション	83

データの質に関する典型的な問題	86
データの内容のエラー	87
ビジネス課題との関連性およびデータの信頼性	89
データの質の定量的評価	91
データ抽出とデータの質——よくあるエラーとそれを避ける方法	92
データ抽出	93
データの妥当性を確かめるための手順	96
派生データ (derived data)	96
データ抽出のまとめ	97
データ入力およびデータ生成がデータの質に与える影響	97
第6章 変数の選択と因子の推定	99
イントロダクション	99
利用可能なデータの選定	100
変数の統計的評価	103
相関	104
因子分析	107
データフュージョン	108
データから変数を選択するアプローチのまとめ	108
望ましい結果を得るための変数選択	109
ビジネス課題に応じて説明変数を評価し選択する統計的手法	109
顧客セグメンテーション	114
変数選択——あらためて分析をやり直す	115
顧客セグメンテーションの最終的なモデル	117
本節のまとめ	120
変数選択に用いるデータマイニングの手法	121
ルールインダクション	121
ニューラルネットワーク	123
クラスタリング	123

パッケージ化されたソリューション	124
オープンソースソフトウェアの利用	124
変数の前選択	125
FAMS (詐欺検出システム)	125
第7章 サンプルングとパーティショニング	129
イントロダクション	129
データを減らすためのサンプルング	130
一定の基準に従ってデータをパーティショニングする	137
サンプルングに伴う問題	141
ビッグデータとサンプルング	142
第8章 データ分析	145
イントロダクション	145
可視化	146
連関	148
クラスタリングとセグメンテーション	149
セグメンテーションと可視化	151
トランザクションデータの分析	157
時系列データの分析	158
データ分析を行う上での典型的なミス	162
第9章 データモデリング	167
イントロダクション	167
モデリングの概念および問題点	167
教師あり学習と教師なし学習	168
クロスバリデーション	169
モデリングの結果を評価する	170

ニューラルネットワーク	173
教師あり学習のニューラルネットワーク	173
クラスタリングを目的としたニューラルネットワーク	176
分類：ルールインダクション	177
ID3 アルゴリズム	179
C4.5 アルゴリズム	180
古典的統計モデル	182
回帰モデル	183
回帰モデルのまとめ	185
k 平均法	185
予測モデル構築におけるその他の手法	186
モデルをデータに適用する	188
「What-IF」を用いたシミュレーションモデル	189
モデリングについてのまとめ	190

第 10 章 システムの開発——クエリレポーティングから EIS

およびエキスパートシステムまで 193

イントロダクション	193
クエリとレポート生成	193
クエリとレポーティングシステム	197
エグゼグティブインフォメーションシステム	198
EIS	201
エキスパートシステム	202
事例ベースシステム	204
まとめ	205

第 11 章 テキストマイニング 207

テキストマイニングの基礎	207
--------------	-----

高度なテキストマイニング	209
キーワードの定義と情報検索	210
個人情報の識別	211
文章抽出	212
情報検索の概念	213
ソーシャルメディアを対象にした感情分析	214
商用テキストマイニングツール	217

第12章 リレーショナルデータベースと連携した

データマイニング

219

イントロダクション	219
データウェアハウスとデータマート	221
データマイニングのためのファイルとテーブルの作成	226

第13章 CRM分析

237

イントロダクション	237
CRMの手法とデータ収集	237
カスタマーライフサイクル	238
リテールバンキングでのCRMの例	241
CRMシステムの統合	243
CRMアプリケーションソフトウェア	244
顧客満足度	245
CRMアプリケーションの使用例	246

第14章 インターネット上のデータを分析する1

——ウェブサイト分析とインターネット検索

255

イントロダクション	255
ウェブサイト訪問者の行動履歴分析	257

Cookie——ユーザ行動のトラッキングと情報の蓄積	260
アクセス解析ソフトウェア	260
インターネット上におけるマーケットセンチメント情報の検索と 統合	261
ウェブクローラとウェブスクレイパー	265
まとめ	269

第15章 インターネット上のデータを分析する2

——検索体験の最適化 271

イントロダクション	271
インターネットとインターネット検索	271
ウェブの構造と検索エンジンにおけるランキングの仕組み	272
インターネット検索のタイプ	276
検索ログのデータマイニング	279
検索行動の表現：クエリセッション	279
検索体験の質の定義	282
検索体験データに関するデータマイニング	284
まとめ	285

第16章 インターネット上のデータを分析する3

——オンラインソーシャルネットワーク分析 287

イントロダクション	287
オンラインソーシャルネットワークの分析	287
グラフ理論における指標	291
グラフデータに用いるデータ形式	293
グラフの可視化と解釈	299
ソーシャルネットワーク分析ツール	305
まとめ	307

第17章 インターネット上のデータを分析する4	
——検索トレンドの時系列変化をつかむ	309
イントロダクション	309
検索トレンドの時系列分析	309
Google Trends——トレンドパターンの分類	311
検索トレンドデータへのデータマイニングの適用	315
トレンドを表現するための説明因子	315
データ抽出と前処理	317
トレンドのクラスタリングと予測モデル	318
まとめ	321
第18章 データにおけるプライバシーと匿名化技術	323
イントロダクション	323
主要なアプリケーションとデータプライバシー	325
法的側面——責任と制限	328
プライバシー保護データパブリッシング	329
プライバシーの概念	329
匿名化技術	332
ドキュメントのサニタイズ	335
第19章 ビジネスデータ分析のための環境整備	341
イントロダクション	341
統合ビジネスデータ分析ツール	341
ビジネスデータ分析のためのアドホック/低コスト環境の構築	347
第20章 おわりに	355

付録 ケーススタディ	357
ケーススタディ 1： 保険会社における顧客ロイヤリティ	357
ケーススタディ 2： リテールバンクにおけるクロスセル	370
ケーススタディ 3： テレビ番組の視聴予測	389
用語集	399
参考文献	405
索引	409