

目 次

第1章 序 論	1
1.1 戦略的データマイニング	1
1.2 戦略的データマイニングの流れ	3
第2章 データ加工	5
2.1 データ加工とは	5
2.2 条件にあったデータを抽出する	5
2.2.1 1つの条件での抽出	7
2.2.2 複数条件での抽出	8
2.2.3 IN条件での抽出	8
2.3 データを並び替える	9
2.4 複数のデータを結合する	10
2.4.1 縦に結合	10
2.4.2 横に結合	11
2.5 ピボットテーブルのような操作を行う	14
2.5.1 縦長のデータに変換する	14
2.5.2 横長のデータに変換する	16
2.6 複雑な集約処理を行う	19
2.6.1 集約したデータを作成する	19
2.6.2 集約結果をもとのデータに追加する	20
2.7 リストからデータフレームを作る	20
2.8 その他細かい処理	22

vi 目次

2.8.1	列名の変更	22
2.8.2	値の置換	23
2.8.3	値の丸め込み	24
2.8.4	分類軸ごとにファイル出力	24
第3章	データ可視化	25
3.1	ggplot2 パッケージ	25
3.2	棒グラフ	29
3.2.1	基本的な棒グラフの作成	29
3.2.2	背景色を白にする	30
3.2.3	グラフタイトルとラベルの設定	30
3.2.4	数値のフォーマットの変更	32
3.2.5	棒の上に値を表示する	33
3.2.6	積み上げ棒グラフの作成	35
3.2.7	色セットの変更	35
3.2.8	棒グラフに枠をつける	37
3.2.9	100%積み上げグラフの作成	38
3.2.10	棒グラフを横に並べる	38
3.3	折れ線グラフ	40
3.3.1	基本的な使い方	40
3.3.2	y 軸の範囲を指定する	41
3.3.3	折れ線グラフに点を追加する	42
3.3.4	折れ線グラフをグループ化する	43
3.4	散布図	45
3.4.1	基本的な使い方	46
3.4.2	散布図をグループ化する	47
3.4.3	回帰直線の追加	47
3.4.4	回帰直線のグループ化	48
3.5	ヒストグラム	50
3.5.1	基本的な使い方	50
3.5.2	棒に枠をつけて色を白にする	50
3.5.3	ビン幅を変更する	51
3.5.4	ヒストグラムをグループ化する	51

3.6	箱ひげ図	53
3.6.1	基本的な使い方	53
3.6.2	箱ひげ図をグループ化する	54
第4章	問題発見	56
4.1	ビジネスにおける問題解決フロー	56
4.2	問題とは	57
4.2.1	現状とあるべき姿のギャップの構造を考える	57
4.2.2	現状とあるべき姿のギャップの構造を調べる	57
4.3	基礎集計で問題を見つける	59
4.3.1	インパクトを比較する	59
4.3.2	分解して比較する	64
4.3.3	トレンドを見る	70
4.4	データマイニングで問題を見つける	78
4.4.1	インパクトの大きい要因と切り口を見つける（決定木分析）	79
第5章	問題解決：意思決定支援	86
5.1	グループ化してターゲットを決める	88
5.1.1	主成分分析	89
5.1.2	クラスタリング	92
5.2	知覚マップを描いてポジショニングを決める	98
5.2.1	コレスポンデンス分析	98
5.2.2	多次元尺度法	104
5.2.3	選好回帰分析	109
5.3	因果関係を把握してアクションを決める	114
5.3.1	因果関係の把握	114
5.3.2	相関分析	118
5.3.3	回帰分析	122
5.3.4	重回帰分析	126
5.3.5	弾力性分析	131
5.3.6	ロジスティック回帰分析	134
5.4	パターンやルールを見つけて施策を決める	140
5.4.1	アソシエーション分析	140

第6章 問題解決：自動化・最適化	154
6.1 自動化・最適化とは	154
6.1.1 予測モデル	154
6.1.2 未知データに対するモデルの評価	155
6.2 ケーススタディ：需要予測を自動化する	156
6.3 モデル構築の可能性を確認する	157
6.3.1 規則性とまとまりの検証	157
6.4 モデルを構築するためのデータセットを作成する	160
6.5 1つのデータセットで予測モデルを構築して比較する	161
6.5.1 線形重回帰モデル	162
6.5.2 ランダムフォレスト	165
6.5.3 SVM 回帰（線形カーネル，多項式カーネル，ガウジアンカーネル）	166
6.5.4 モデルの比較	168
6.6 予測モデルの自動選択を実装する	170
6.7 予測の精度をあげるには	173
第7章 効果測定	175
7.1 基本的な考え方	175
7.2 ランダムなグループ間で比較する	176
7.2.1 偏りの確認	177
7.2.2 A/Bテスト	177
7.2.3 多変量テスト	181
7.3 非ランダムなグループ間で比較する	185
7.3.1 層別解析	185
7.3.2 傾向スコア解析	192
付録A R/RStudioの導入	199
A.1 Rの導入	199
A.1.1 Windows	199
A.1.2 Max OS X	204
A.1.3 Linux	204
A.2 RStudio	206
A.2.1 RStudioとは	206
A.2.2 RStudioのインストール	206

A.2.3 RStudioの基本操作	209
A.3 Rの基本操作	212
A.3.1 算術演算子	212
A.3.2 比較演算子	213
A.3.3 論理演算子	214
A.3.4 集計関数	215
A.3.5 データの読み込み	216
A.3.6 データフレームの操作	217
A.3.7 パッケージの利用	218
索引	219