

データマイニングによる顧客スコアリング

(申請中) 筑波大学 *後藤正輝 GOTO Masateru
筑波大学 村山和穂 MURAYAMA Kazuho
筑波大学 門間公志 MONMA Koji
01105930 筑波大学 香田正人 KODA Masato

1 はじめに

データマイニングは大規模データベースから未知の有効な情報を抽出するプロセスである。また顧客スコアリングとは過去の購買履歴データを基にして、購入可能性の高い順に顧客をランク付けするデータマイニングの一分野である。

本研究では企業から提供を受けた取引履歴データに対して複数のデータマイニング手法を適用し、顧客スコアリングモデルを作成した。また作成したモデルを基にテストマーケティングを行い、その有効性を検証した。

2 利用データ

本研究で利用したデータはある衣料・雑貨販売会社の通信販売履歴データである。原データは取引IDをキーにもつ販売履歴データに、商品属性・顧客属性に関する情報を付加したものであるが、本研究の目的より顧客IDをキーにした分析用データを作成した。

分析用データは12,242名の顧客に対する1998年10月から2001年5月までの取引履歴であるが、カタログ発送スケジュールに従い、98年度秋冬期(以下98aw)から01年度春夏期(以下01ss)までの半年毎に顧客の購買行動を捉えてモデルの説明変数を作成した。テストマーケティングでは01年度秋冬期の購入者を予測した。

3 スコアリングモデルの作成

3.1 問題のモデル化

顧客スコアリングは本質的には顧客を将来の購入有無で分類するクラス判別問題であるが、いかに巧妙な説明変数を用いても2クラスの判別境界を定めることはできない。そのため予測期間中の取引有無という離散値を教師信号として、購入確率という連続値を出力する回帰モデルを作成して顧客をランク付けする。

3.2 利用アルゴリズム

スコアリングモデルの作成にはIBM DB2 Intelligent Miner for Data V6.1を使用した。利用したアルゴリズムはバックプロパゲーションネットワーク、ラジアルベシス関数ネットワーク、回帰木(以下それぞれBPN,RBF,RT)の3種である。ここに、RBFの基底関数の決定はニューラルクラスタリングにより行われ、RTのノード分割規準はGiniインデックスによる。

3.3 モデルの評価

本研究の目的は現実の業務への適用において高い予測精度を発揮するモデルを選定することである。したがって、00awの購入有無をそれ以前の購買行動から予測するモデルを訓練し、それを入力期間を1期間ずらして01ssの予測

に適用しテストする。以上を繰返し最も高い予測精度をもつモデルを選択する。

予測全体の当てはまりを評価するためにリフト図を用いる。リフト図は顧客をランクに従い降順にソートし、反応者数の累積値をプロットしたものである。また、予測スコア上位 $x\%$ の顧客群に対する精度の定量的な評価にはリフト率を用いる。

$$\text{リフト率}(x) = \frac{\text{上位 } x\% \text{ の顧客の反応率}}{\text{全顧客についての反応率}}$$

4 結果

予測精度の評価は利用アルゴリズム、説明変数の種類、説明変数の次元の3点について比較を行った。その結果アルゴリズム別ではBPNが、また説明変数については4期間の購入回数のみを入力としたものが高い予測精度を示した。

得られた最適モデルにしたがって、予測上位1,000名の顧客を選定した。11月時点でのテストマーケティングの結果は実際に購入した顧客が195名であり、反応率は19.5%であった。全顧客についての反応率は5.2%であり、上位1,000名についてのリフト率は3.74であるので、テストマーケティングにおいてもモデル選択段階とほぼ同等の予測精度を得ることができた。

最後になりますが、データ提供をいただいたI社に感謝いたします。

参考文献

- [1] Bigus, Joseph P. (1996):Data mining with neural networks. McGraw-Hill Companies.
- [2] Peter Cabena, Pablo Hadjinian, Rolf Stadler, Jaap Verhees, Alessandro Zanasi (1997):Discovering Data Mining - From Concept to Implementation. Prentice Hall

表 1: 分析用データ

属性	データの内容
key	[顧客ID]
in1-4	期別購入回数
in5-8	期別購入金額
in9-12	期別購入商品数
in13	購入回数合計
in14	購入金額合計
in15	購入商品数合計
in16	購入単価
in17	注文あたり商品数
in18	初回購入時期
out	予測期間の購入有無

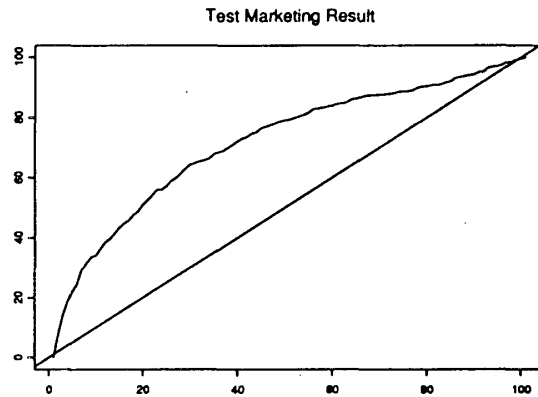


図 1: リフト図。横軸はランクの高い順にソートされた顧客一覧、縦軸は累積購入者割合。

表 2: 利用アルゴリズムによる比較

	BPN	RBF	RT
リフト率	3.88	3.51	2.58

表 3: 説明変数の種類による比較

	回数	金額	商品数
リフト率	3.88	3.51	2.96

表 4: 説明変数の次元による比較

	4次元	6次元	18次元
リフト率	3.88	3.74	3.65