

移動体通信業界へのデータマイニングの適用

5 K - 9

鮎川 江里香*

森田 豊久*

森繁 誠美†

* (株) 日立製作所 システム開発研究所

† (株) デジタルツーカー中国

1. はじめに

計算機の高速化、ディスクの大容量化、ネットワークのオープン化などにより、多くの分野・業種で大規模なデータの蓄積が可能となるにつれて、これらの大量な蓄積データから有用な知識を発掘するデータマイニングが注目を集めるようになってきた。

例えば、近年急速な成長を遂げつつある移動体通信業界では、加入者数の増大が著しく、各事業者内に蓄積される情報量も膨大なものになりつつある。その一方で、加入者の大半が比較的短期間のうちに解約してしまい、十分な利益を獲得しきれないという深刻な問題も抱えている。そこで、同業界では日々蓄積される大量の顧客データを活用し、解約顧客の傾向をさまざまな角度から分析することで、解約要因の特定や解約率の低減のための施策を練りたいと考えている。

本稿では、上述の移動体通信業界へのデータマイニング適用事例として、大量データ中に存在する規則性や因果関係を If-Then ルールの形で抽出する特徴的ルール生成方式[1]を用いた顧客の特性分析と、その評価結果について報告する。

2. 特徴的ルール生成方式

特徴的ルール生成方式(CHRIS:Characteristic Rule Induction by Subspace Search)では、大量データ中の規則性や因果関係を、「もし～ならば…である」という If-Then ルールの形で抽出する。各々のルールは、変数名(年齢など)と、その値をあらかじめいくつかのカテゴリに変換(カテゴリ化)して得られるカテゴリ値(小, 中, 大など)との組を用いて表現される。生成されるルールの形式は以下の通り。

```
IF
  X1 = x1 & X2 = x2 & … & Xn = xn
THEN
  Y = y
```

ここで、 X_i は説明変数名、 x_i は変数 X_i のカテゴリ名である。同様に Y は被説明変数名、 y は変数 Y のカテゴリ名である。「 X_i が x_i 」という組を条件節と呼ぶ。また、「IF A THEN B」の A を条件部、B を結論部と呼ぶ。

それぞれのルールには、評価尺度[1]と呼ぶ実数値を割り当て、評価尺度の高いルールほどルールの価値が高いものとみなす。本方式では、あらゆる条件節の組み合わせからなるルールに対し評価尺度の算出を行い、評価の高いルールのみを選択する処理を実行する。

3. 移動体通信業界における顧客特性分析

今回は、解約顧客の中でも、特に料金滞納などの理由で事業者側が強制的に解約を行う、いわゆる「強制解約」顧客の特性分析を試みた。

(1) データ

強制解約顧客を含む全顧客の属性データ、及び利用データを使用した。ここで、属性データとは、「年齢」、「性別」や「料金プラン」などの属性情報、及びその変更履歴である。また、利用データとは、「通話回数」や「通話時間」など、日々のサービス利用状況に関する実績情報である。

属性データ、及び利用データをファイル結合して得られる分析用データを図1に示す。分析用データには、それぞれのレコードが強制解約顧客に相当する(「契約状態」=「強制解約」)か否(「契約状態」=「継続」など)かを識別するための変数「契約状態」も含まれている。

レコード 番号	属性データ			利用データ			契約 状態
	年齢	性別	...	通話 回数	通話 時間	...	
1	37	男性	...	7	0:46	...	継続
2	23	男性	...	36	1:02	...	強制解約
3	56	女性	...	20	1:52	...	強制解約
:	:	:	...	:	:	...	:

図1 分析用データ

(2) 方法

If-Then ルール (IF A THEN B) の結論部 B に強制解約顧客 (「契約状態」=「強制解約」) を指定し、強制解約顧客をできるだけうまく説明する条件節の組 (すなわち、条件部 A) を探索した。

(3) 結果

(2) により、強制解約顧客の特性を示すさまざまなルールが得られた。その一例を図2に示す。

IF	1 通話当たりの平均着信時間 = 小 & 発信時間5~10分の頻度 = 大 & 年齢 = 大
THEN	契約状態 = 強制解約

図2 ルールの一例

4. ルールの評価

3. で得られた上位 50 個のルールについて、その評価を行った。

図3にルールのリフト[2]を示す。図3のX軸は、If-Then ルール (IF A THEN B) の条件部 A を満たすレコードの割合 $P(A)$ を表す。Y軸は、If-Then ルールの結論部 B を満たすレコードの割合 $P(B)$ に対し、条件部 A と結論部 B をともに満たすレコードの割合 $P(A\&B)$ の比 $P(A\&B)/P(B)$ を表す。また、直線 $y = x$ は、ある一定の割合のレコードを無作為に抽出した場合に、それらのレコードに含まれる強制解約顧客の全強制解約顧客に占める割合を示している。ここで、無作為に抽出されたレコードに含まれる強制解約顧客の全強制解約顧客に占

める割合 (α) と、ルールの条件部により絞り込まれたレコードに含まれる強制解約顧客の全強制解約顧客に占める割合 (β) の比 (β / α) を「ルールのリフト」と呼ぶ。図3に α , 及び β の一例を示す。

図3より、リフト10以上のルールも複数確認され、3. で得られたルールが強制解約顧客を効果的に絞り込んでいることが明らかとなった。

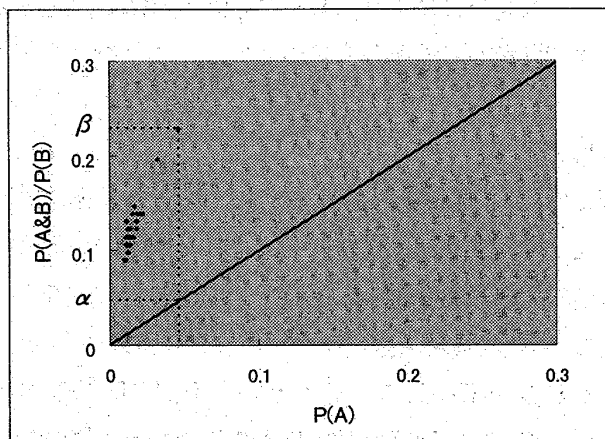


図3 ルールのリフト

5. おわりに

本稿では、移動体通信業界へのデータマイニング適用事例として、特徴的ルール生成方式を用いた顧客の特性分析と、その評価を行った。その結果、上記分析方法が強制解約顧客の絞り込みに対し有効であることを確認できた。

参考文献

[1] 芦田仁史ほか: データマイニングにおける特徴的ルール生成方式, 情報処理学会第50回全国大会, 3-19 (1995)
 [2] Berry, M.J.A. and Linoff, G.: Data Mining Techniques for Marketing, Sales and Customer Support, John Wiley & Sons (1997)