

データマイニングによる金融データ解析

平山 克己, 田村 亮二

1. はじめに

OR学会のマーケティングエンジニアリング研究会、2000年度データ解析コンペティションに参加し、発表した内容をこの場をお借りしてご紹介いたします。今回のコンペティションの対象データは某銀行の貴重なもので、ペイオフや超低金利時代など激しく変化する環境下での預金者の意識を探るアンケート（意識データ）と各預金者のトランザクションデータ（行動データ）を組合せたものでした。つまり、「預金者の意識データと行動データを如何に解析するか？」というやりがいのあるテーマでした。我々はデータマイニングツール SAS Enterprise を用いて、この複雑なデータ解析に挑戦しました。

データマイニングについては、本誌 Vol. 43, No. 12 に特集[1]があるので、興味のある方はそちらを参照してください。文献としては文献[2]が参考になると思います。

2. 分析の目的

データ分析を行う際、注意しなければならないのは多角的な視点から客観的にデータを分析することだと思います。実際、分析に熱中すると途中で何をやっているかが判らなくなってしまう、時間と労力だけがかかってしまうという経験を私は何度も味わいました。そこで、今回は「銀行経営者の立場に立ち、どのような情報をデータから読み取ることができるか？」という分析の方針を立てました。金融ビッグバンを経て、銀行の競争が激化している中で銀行経営者は様々な課題を抱えていると思いますが、競争に勝つためには、大きく分けて、2つのアプローチがあるのではないかと

と考えました。

1つ目は取引高拡大のためのアプローチであり、2つ目は、コスト削減のためのアプローチであると考えました。しかし、今回のデータにはコスト削減に必要なコスト情報が含まれていないので、今回は1つ目の取引拡大のアプローチを中心に考えていくことにしました。これは、リテールバンキングと呼ばれている“個人顧客の囲い込み”、別の言い方をすればCRM (Customer Relationship Management) であるともいえます。そして、取引拡大のための具体的なアプローチとして、我々は次の4点に絞って考えていくことにしました。

1. 優良顧客へのサービス拡充
2. 既存顧客つなぎとめ
3. 新規顧客チャネルの発見
4. 魅力ある商品開発（商品構成の見直し）

1~3. はどれも「顧客を上手く分類し、それぞれに対してどのような戦略をとるか？」に関わっているのではないかと考えました。そして、4. は商品組合せと商品推移の把握につながると考えました。つまり、顧客の顔（属性）が判らなければ、具体的な経営アクションを取れないのではないかと考えたのです。ダイレクトメール1通出すにもコストがかかり、顧客全員に出していたのでは膨大なコストがかかってしまうので、ヒット率の高い顧客を見つけようというデータマイニングの考え方を使って顧客を分類することにしました。また、どのような顧客が優良顧客で、どのような顧客が離れていっているのかを分析することはマーケティングには欠かせない分析であるはずですが、しかし、顧客の意識は時と共に変化します。昨日の優良顧客が、明日は非優良顧客になってしまう可能性もあります。そこで、我々は次の2つの視点からデータ分析を進めて行くことにしました。

【視点1】アンケート+取引データによる静的分析

- 分析1：アンケートによる顧客分類
- 分析2：決定木による顧客属性の分析

ひらやま かすみ

北九州市立大学 経済学部

〒802-8577 北九州市小倉南区北方4-2-1

たむら りょうじ

㈱住友金属システムソリューションズ

〒110-0008 東京都台東区池之端1-2-18

●分析3：金融商品のバスケット分析

【視点2】取引データ変動パターンによる動的分析

●分析4：変動パターン分析

●分析5：解約者分析

●分析6：時系列を考慮したバスケット分析

各分析について、次節で詳細に説明して行きます。

3. 分析内容

まず、分析を進めて行く前にデータがどのようなものを事前分析したものが図1です。この図は、サンプルデータの中から99年7月のトランザクションデータをもとに、顧客別の総資産を求め総資産額の多い顧客を左から並べた棒グラフ（左軸）とその比率を累積した曲線（右軸）です。また、棒グラフ中、縦に線が入ったところは解約した顧客です。この図から、1000万円以上の資産を持っているユーザは983人で、全サンプル数の約1/3を占めていました。また、全サンプル数の1/3の顧客の資産総額はサンプルデータの資産総額250億の59%を占めていることが判りました。

ここで、1000万円というのはペイオフが実施された場合、1つの目安となる金額です。なぜなら、ペイオフが実施された場合、1000万円以上の資産を持っている顧客は資産を分散させる可能性があるからです。つまり、図中の△ABCに含まれている預金などがどこかに行ってしまう可能性があるのです。もちろん、逆に△ADEに含まれている顧客が新たに預金を増やしてくれる可能性もあるわけですが…。したがって、ペイオフが実施された場合、どの客が優良顧客になるかは、この1時点でのトランザクションデータからは判らないわけです。

3.1 分析1：アンケートによる顧客分類

郵便局の定額貯金の利率が銀行に比べ著しく良かつ

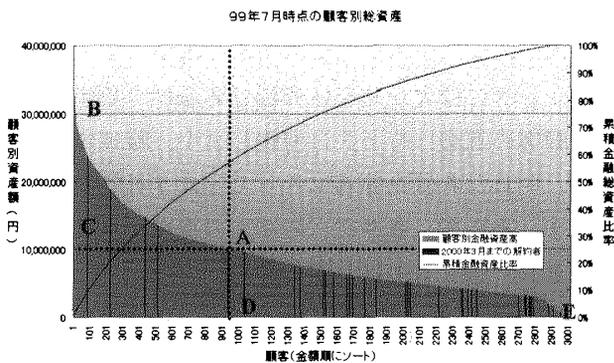


図1 事前分析

た時期に定額貯金に預金した人が2001年7月頃満期となり、約10兆円規模の貯金が市場に溢れるというニュースがありました。このニュースから郵便局の定額貯金者を獲得するのが得策であると考えました。アンケートデータQ1とQ3から、郵便局に預金している顧客の数を見ると、全サンプル数の約3/4の顧客が郵便局に貯金していることがわかりました。そこで、我々はアンケートデータから当行以外にも預金している客と当行だけに資産を預けている顧客を分類できないかどうか検討してみました。ここで、前者を流動的な顧客とし、後者を固定的な顧客と呼ぶことにします。

当初、アンケート項目のQ2“いくつの「都市銀行」とお取引がありますか？”という質問項目から流動的な顧客と固定的な顧客を判断できると考えましたが、回答内容に矛盾が多く使用するのが困難でした。

そこで、次の複数の質問項目から固定的な顧客と流動的な顧客を分類することにしました。

●Q1：現在、お取引のある金融機関に○印をつけてください（複数回答可）。

- 1 都市銀行, 2 地方銀行, 3 信用金庫, 信用組合,
- 4 信託銀行, 長信銀, 5 郵便局, 6 証券会社, 7 外資系金融機関, 8 その他

●Q3：現在どのような金融商品をお持ちですか？

- 1 普通預金（銀行）, 2 貯蓄預金（銀行）, 3 定期預金（銀行）, 4 郵便貯金（郵便局）, 5 MMF, 中国ファンド, 6 公社債投信, 7 株式投信, 8 株式（従業員持ち株会は除く）・社債, 9 外貨預金, 10 保険会社の貯蓄型保険, 11 財形・社内預金・従業員持ち株会, 12 その他

●Q5：以下の金融機関の中で、預けている金額が一番多いのはどちらですか？

- 1 都市銀行, 2 地方銀行, 3 信用金庫, 信用組合,
- 4 信託銀行

●Q9：以下の金融機関のうち、普段のご利用（入出金）が一番多いのはどちらですか？

- 1 都市銀行, 2 地方銀行, 3 信用金庫, 信用組合,
- 4 信託銀行

上のようなアンケート項目において、Q1で1のみに○、かつQ3で4以外に○、かつQ5とQ9で1に○をつけた顧客を固定的な顧客とし、それ以外を流動的な顧客としました。ただし、総資産が0または欠損のレコードは除外しています。

顧客の分類結果を図2に示します。この図は図1の直線ADで左右に分類した顧客を更に固定的な顧客

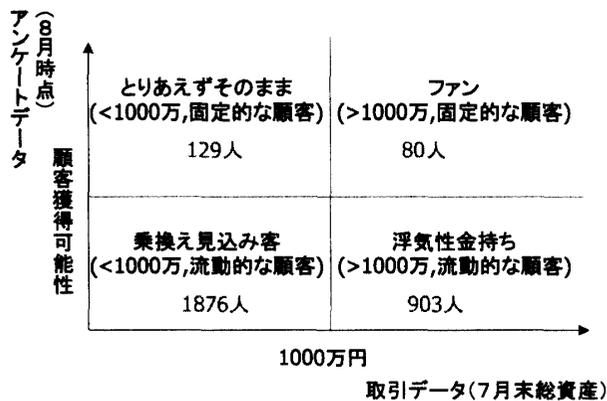


図2 顧客セグメンテーション

と流動的な顧客に分類したことになります。つまり、99年7月時点での顧客の状態と99年8月時点での将来流動的な顧客から固定的な顧客に変わってくれる見込み客で分類したことになります。

図2の右上に分類される顧客は固定的な顧客、かつ99年7月現在で当行に1000万円以上預金している顧客なのでファンセグメントとしました。右下に分類される顧客は流動的な顧客、かつ1000万円以上預金している顧客層なので浮気性金持ちセグメントとしました。次に、左上は固定的な顧客、かつ1000万円以下の預金者なのでとりあえずそのままセグメントとしました。最後に、左下は流動的な顧客、かつ1000万円以下の預金者なので、乗り換え見込み客セグメントとしました。それぞれのセグメントの預金者数は80, 903, 129, 1876人となりました。我々は取引拡大のための戦略上、ファンセグメントが最も重要で、引き続き、乗り換え見込み客、浮気性金持ち、とりあえずそのままの順に優先順位をつけました。

3.2 分析2：決定木による顧客属性の分析

ここでは、「分析1で分類した各顧客層間がどのような意識の違いがあるのか?」について、データマイニング手法の1つである決定木を使って分析しました。その結果を図3に示します。この決定木は顧客セグメンテーションに使ったアンケート項目以外をすべて説明変数として、ターゲット（目的変数）を顧客セグメントとしました。アルゴリズムはCHAIDを選択しました。ここでは、我々が分類した顧客セグメントはどのようなアンケート項目によって分類できるかを分析したものです。

図中の四角(葉)内の数字1, 2, 3, 4はそれぞれファン、乗り換え見込み客、浮気性金持ち、とりあえずそのままセグメントを示しています。この図から、ア

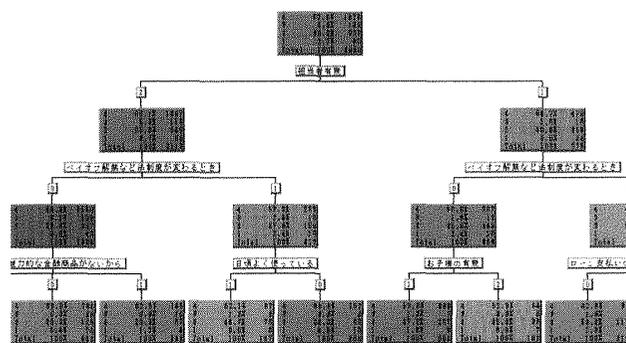


図3 顧客セグメンテーション別決定木分析

ンケート項目Q8“その金融機関には、自宅や職場にきてもらったり、電話で相談できる担当者がいますか?”という項目が決定木の最初の分岐となっていることから、Q8が顧客セグメントを分類する上で最も有意な説明変数であるといえます。

つまり、預金や出金の際にわざわざ出向くのではなく、デリバリーサービスが重要なファクターとなっているという解釈が可能で、更に、この項目によって左右に枝分かれした葉の中を見てみると、浮気性金持ちの比率が左右で大きく異なることが判ります。このことから、浮気性金持ちは出不精なので預金出金のデリバリーサービスが預金先銀行を決める大きなファクターになっていることが判ります。

次に現れた項目はQ2“あなたのお宅では、貯蓄の運用方法を考え直すのは、どのようなタイミングですか? (複数回答可)”という質問に対して、10“ペイオフ解禁など法制度が変わるとき”に○をつけたか否かが有意な差となるということです。これは右の葉も左の葉も同様の項目が出てきていることから、担当者の有無に限らず我々の顧客セグメンテーションに対して有意な説明変数であるといえます。

以上の結果から、図2で現在の預金状態と顧客獲得可能性という2次元によって分類した顧客セグメントは、担当者の有無(デリバリーサービス)と法制度の変化が起こるタイミング(外部環境の変化)という別の次元に変換することができると考えられます。この変換に可逆性があるとするれば、「デリバリーサービスを向上し、外部環境の変化のタイミングを見計らえば、流動的な顧客を固定的な顧客に変えることができる」という解釈も可能であると考えられます。

3.3 分析3：金融商品のバスケット分析

分析2では、顧客を獲得するため重要な要因がサービスの形態とタイミングであることが解りました。そ

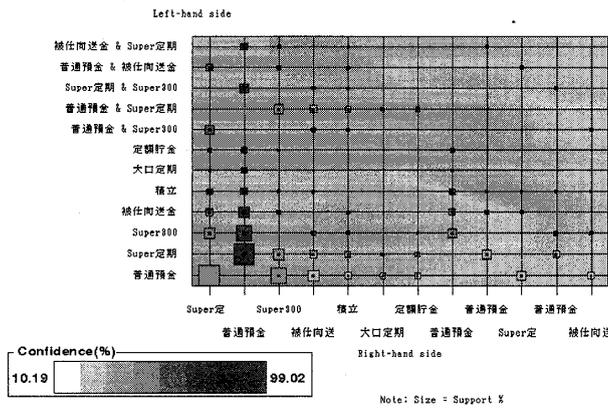


図4 金融商品バスケット分析

れでは、次に「どのように商品を組合せて提供すればよいか？」について分析するために、バスケット分析を使いました。その結果を図4に示します。この図は縦軸に並んでいる商品と同時に横軸に並んでいる商品をもっている確率を交点の正方形の大ききで示しており、正方形が大きいほどこの確率（指示度）が高くなっています。また、正方形の色が濃いほど縦軸に並んでいる商品を持っている顧客のなかで横軸に並んでいる商品を持っている確率（信頼度）が高いことを示しています。

例えば、図左下の縦軸が普通預金、横軸がSuper定期の交点にある正方形と、その右上にある縦軸がSuper定期、横軸が普通預金の交点にある正方形の大ききは同じなので指示度としては同じですが、右上にある正方形の方が色は濃いので、普通預金を持っている顧客の中ではSuper定期を持っている顧客は相対的に多く、Super定期を持っている顧客は、当然、普通預金を持っていることを示しています。この図からも判るように、この分析からは当たり前のことしか判りませんでした。

3.4 分析4：変動パターン分析

ここでは、トランザクションデータを使い、預金者の行動を分析してみました。図5は顧客別の普通預金と総資産に対して、99年3月から1年分の変動を見たものと、四半期ごとの総資産を見たものです。それぞれ、前月に比べ増額していれば+、減額していれば-、変動していなければ0と表記し、変動パターンの頻度について分析したものです。毎月の変動パターンを全列挙すれば3の12乗で53万通りという膨大なパターンが出てくるはずですが、今回の分析では普通預金で1627通り、総資産で1689通りのパターンが出現しました。

出現パターン 頻度		出現パターン 頻度		出現パターン 頻度		
34567890DN123 (月)		34567890DN123 (月)		369N3 (月)		
1	000000000000	387	1 ++++++	159	1 +++	480
2	+++++	140	2 ++++++	126	2 +-+	290
3	0000+0000+0	103	3 000000000000	71	3 +++	194
4	-----	37	4 -----	31	4 +++	191
5	+++++	27	5 0000+0000+0	27	5 ---	162
6	-----	23	6 -----	24	6 +-+	160
7	+++++	23	7 00000000+000	19	7 +-+	132
8	+++++	22	8 ++++++	18	8 +++	131
9	-----	17	9 ++++++	17	9 +++	125
10	+++++	13	10 ++++++	16	10 +-+	98
11	+++++	13	11 ++++++	15	11 +-+	93
12	+++++	12	12 ++++++	13	12 +-+	86
13	+000+0000+0	11	13 ++++++	13	13 ---	81
14	+++++	11	14 ++++++	13	14 +-+	73
15	+++++	11	15 -+++++	12	15 0000	70

普通預金 総資産 4半期総資産

図5 変動パターン分析

毎月の変動にするとパターン数が多く1パターンあたりの頻度が小さくなるので、四半期毎の変動パターンも見てみましたが、面白い結果は現れませんでした。この図の特徴としては、まず普通預金の変動パターンで3番目に頻度の高いパターンでボーナス時期だけ増額している顧客が103人ほどいるということです。また、普通預金では毎月増額しているパターンは27人しかいなかったにもかかわらず、総資産で見ると159人もいるということです。この159人の中には積み立て貯金をしている人も含まれています。更に、普通預金では最も頻度の高かった毎月変動額0のパターンが、総資産のパターンで見ると3番目となり、逆転していることです。これらの考察から、金額の大小はあるものの、普通預金以外の金融商品で着実に資産運用している顧客がいるのではないかと仮説を立てました。

それでは「着実に資産運用している顧客はどのような意識をもっており、どのような行動をとっているのでしょうか？」我々は四半期総資産の変動が全て+の顧客を1とし、それ以外の顧客を0として、アンケートデータとトランザクションデータを基に決定木分析を行いました。その結果が図6です。

この図から、年間変動額が約2000円以下の客は四半期総資産の変動が全て+の顧客ではないことが判ります。なんとなく、当たり前の結果のように思えますが、着実に資産運用している客の行動はアンケートデータの項目、つまり意識とはあまり相関がなく、年間の変動額に依存しているということが出来ます。

次に、普通預金以外のどのような商品にどれくらいの額で資産運用しているのかを分析しました。まず、毎月総資産の変動が+の顧客の中から、99年3月から2000年3月までの円貨預金の変動額が高い順に30人ほどピックアップし、リスク性商品（外貨、投信、

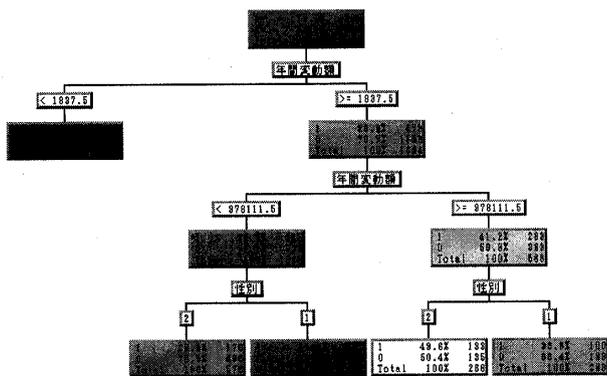


図6 特定変動パターンの決定木

出現パターン	頻度
34567890DN123 (月)	
000-XXXXXXXX	5
0000000-XXXX	5
00-XXXXXXXXXX	4
0000-XXXXXXXX	4
000000000-XX	4
0000000000-X	4
+XXXXXXXXXX	3
++XXXXXXXXXX	2
00000-XXXXX	2
000000-XXXX	2
00000000-XXX	2
+++-----XX	1
++++-----XXXX	1
++-----XXX	1
-----+X	1

図8 解約者の変動パターン分析

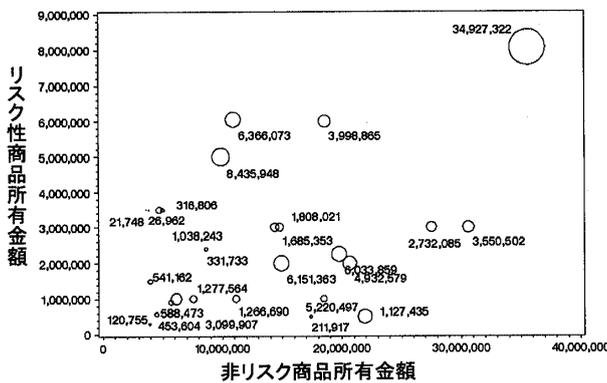


図7 リスク性商品と非リスク性商品のバブル図

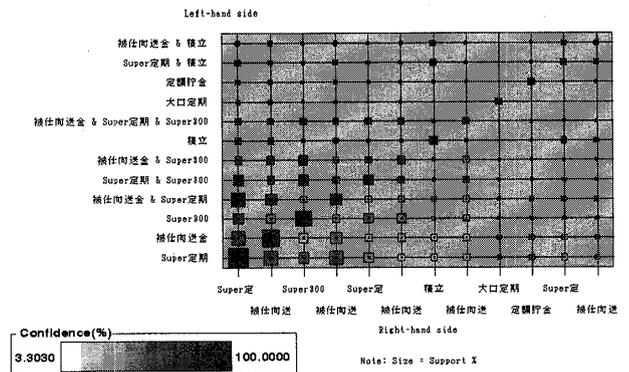


図9 時系列を考慮したバスケット分析

国債、抵当証券)と非リスク性商品(円貨預金)のどちらで運用しているかを調べたものが図7です。

図中の円はピックアップした30人の顧客における、リスク性商品と非リスク性商品の保有状況を示したものです。円の中心点は2000年3月時点でのリスク性商品の預金総額(縦軸)と非リスク性商品の預金総額(横軸)を示しており、円の直径が大きいほど99年3月から2000年3月までに預金した額が大きくなっています。この図から、総資産額に対する非リスク性商品の割合はだいたい5分の1前後であることがわかります。リスクに対する投資額という点から見ても、自身の資産(額は小さいのですが)運用経験からしてみても妥当な割合のように見えます。

3.5 分析5:解約者分析

ここでは、解約者についての毎月の変動パターンを分析してみました。図8は図4と同様に解約を“X”と表記し、変動パターンの頻度を調べたものです。全体のサンプルが少ないためあまり面白くない結果ですが、幾つかの傾向が見て取れます。

1つ目は毎月変動無しが継続し、一気に引出し、解約するパターン“00-”が顕著に見られる。解約する場合は段階的に引出しを行うのではなく、一度に全額

を引き落としている傾向が見てとれます。このことは、解約防止のために“00-”のパターンを注意深く監視しても、時すでに遅しということを示しています。したがって、解約防止のためには“00”というパターンを注意深く監視し、このパターンを示した顧客に対しては何らかの対策を打つことが重要になってくると考えられます。

3.6 分析6:時系列考慮のバスケット分析

図9は図4と同様に、バスケット分析を行ったものです。図4とは違い、99年3月~2000年3月の一年分のトランザクションデータを月毎に見て、各金融商品の有無を考慮し、時系列に展開して分析したものです。この図から、Super定期を保有している顧客は、被仕向送金を行っている確率が高いことが判ります。また、Super定期を保有している顧客は他の金融商品も保有している確率が高いことが判ります。「どの金融商品を積極的に売り込むか?」という問いに対しては、Super定期という回答がこの結果から読み取れます。

4. 分析結果からの提案

これまで述べてきた、6つの分析から次の5つの具

