

# 複数 MBR モデルを用いた与信スコアリング高精度化の実現

1 Y-06

吉川 裕      野崎 賢      安信 千津子  
 (株) 日立製作所      システム開発研究所

## 1. はじめに

近年、金融業界の与信業務において、融資申込顧客やクレジットカード申込顧客の信用度をスコアで表現する与信スコアリングモデルの活用がさかんになっている。従来の与信スコアリングモデルでは、多変量解析やニューラルネットワークによる予測モデルが利用されてきた[1]。しかし、顧客の嗜好変化や社会環境の変化に即したモデルを維持するためにはコストと時間が掛かり、急激に変化する現在のビジネス環境に対応できないという問題があった。

このような中で、データマイニングの一手法である MBR (Memory Based Reasoning) を応用した与信スコアリングモデルが注目を集めている。図 1 に示すように、MBR は予測対象データと既知データとの類似度を利用して予測する手法であり、データを追加・更新するだけ予測対象の変化に追従可能といった特徴がある[2]。

しかしながら、MBR による与信スコアリングにおいても、類似データを探索するデータ（記憶事例）として用いる、「不良」の既知データが十分に得られない場合には、十分な予測精度が得られないことがある。また、既知データを複数のデータセットに分割し、各 MBR モデルの予測結果を検証したところ、あるモデルで正常に識別できる顧客であっても、別のモデルでは誤識別してしまうことがあった。これらの結果は、記憶事例のデータ数が不十分な場合、予測精度がデータ分布の偏りや異常値に影響を受けやすくなるためだと考えられる。

そこで本稿では、上記の問題を解決するために、複数の MBR モデルを用いた与信スコアリングの高精度化方法を提案する。

融資の初期与信への適用例

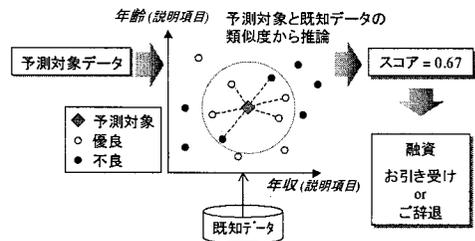


図 1: MBRによる与信スコアリング

## 2. 提案方法

提案方法では、Boosting / Bagging 法を応用し、与信スコアリング高精度化を図る。Boosting / Bagging 法は、精度の低い予測モデルを組合せることで、高精度な予測モデルを得るための手法であり、各予測モデルの予測結果の多数決で最終的な予測を決定する[3]。提案方法では、図 2 に示すように、記憶事例と説明項目が異なる複数の MBR モデルを使用し、各モデルのスコアから与信審査に用いる識別スコアを算出する。

提案方法

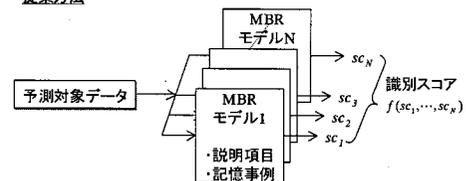


図 2: 複数 MBR モデルを用いた与信スコアリング

本稿では、以下の識別スコア算出式を用いた与信スコアリング方法について、その有効性を検証する。

$$SC1 = \frac{\sum sc_i}{N} \quad \text{式(1)}$$

$$SC2 = \frac{\sum sc_i - \max(sc_i) - \min(sc_i)}{N - 2} \quad \text{式(2)}$$

以下、上記の各式について説明する。

まず、式(1)は全モデルのスコアの平均値を最終的な識別スコアとするものである。平均値を利用することで、異常値の影響を除外する効果があると考えられる。

式(2)は、全モデルのスコアのうち最大値と最小値を除外した平均値を、最終的な識別スコアとするものである。最大値と最小値を除外することで、式(1)よりも異常値の影響を除外する効果があると考えられる。

なお、本稿での与信スコアリングモデルで算出するスコアは0以上1以下の実数値であり、スコアが0に近ければ信用度の高い「優良」顧客であり、1に近ければ信用度の低い「不良」顧客であることを意味する。また、既知データにおける顧客が、「優良」と「不良」のどちらであるかは、延滞や貸倒の回数、金額などに基づいて決定する。

### 3. 実験結果

モデル構築用データとは別に用意した既知データに対して、提案したスコア算出式を用いた識別スコアを算出し、その予測精度を評価した。

図3は、予測精度の評価グラフ(リフト図)である。スコア上位x%に不良のy%が含まれていることを示している。本稿では、モデル数が5の場合の式(1)、式(2)による結果および、比較のため1モデルと無作為抽出の結果を示す。無作為抽出の場合の直線よりも、評価曲線が上にあるほど、審査能力の高いモデルである。

なお、各モデルの記憶事例は、モデル構築用データから無作為抽出により作成している。また、各モデルの説明項目は、1モデルで予測したときに最高精度となるよう、予め選択してある。

図3から、1モデルに比べて式(2)および式(1)ともに精度が向上していることが分かる。本例では、スコア上位20%に相当するスコア以上の顧客に対する融資を自動謝絶する場合を想定しており、このときの不良の識別能力は、式(2)では約14ポイント、式(1)では約10ポイント向上した。

### 4. 考察

以上から、提案したスコア算出方法のうち式(2)ないし式(1)を用いることで、不良の識別能力が改善すると考えられる。特に、式(2)を用いることで、1モデルに対して最も精度向上があった。式(2)においてスコアの最大値と最小値を除外することで、異常値の影響を除外する効果があったと考えられる。

これらの結果から、複数のMBRモデルでの予測を統合して識別スコアを算出することで、異常値に影響されにくい頑健性が強化されることが確認できる。

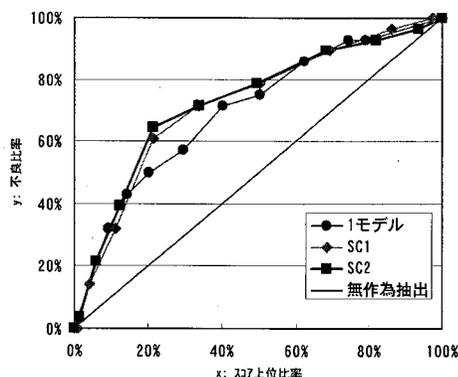


図3: 評価グラフ(リフト図)

### 5. まとめ

複数のMBRモデルを組合せて、与信審査の判定基準となるスコアを算出する与信スコアリング手法を提案した。本方法により、異常値に影響されにくい頑健性の高い与信スコアリングを実現できることを確認した。

### 参考文献

- [1] エリザベス・メイズ, "クレジットスコアリング", シグマベイスキャピタル, 2001
- [2] 松本和宏, 前田一穂, 柳沼義典他, "金融ビジネスユースに適したデータマイニング手法: MBR", 情報処理, Vol.42, No.7, pp.696-701, 2001.
- [3] Yoav Freund, Robert Schapire, "ブースティング入門", 人工知能学会誌, Vol.14, No.5, pp.771-780, 1999.