スコアリングによる与信リスク管理

白川　浩

1. スコアリングとは

経営活動においては，現有資源を有効に活用し企業体質をより強固にするために，日々さまざまな経営意思決定を適切に行う必要がある．このためには直面する意思決定が創造的な解決法を必要とする非定形的意思決定なのか，ある基本的な特徴を捉えることにより類型化可能な定型的意思決定なのかを明確に区別し，適切に対応することが重要となる．すなわち経営資源を有効に利用するには，人材という非常に高価な資源は非定形的意思決定に集中的に投入し，定型的意思決定はマニュアル化により意思決定の質の向上を経済的に実現する必要があるのである．

この定型的意思決定問題の具体的特徴を表す与件要素の抽出と，判断の優先順位を求める仕組みがスコアリングと呼ばれる技術である．スコアリング技術を経営活動の便利な道具として広く利用するには，基本的に3つの課題が存在する．すなわち，(1)意思決定結果の質の向上を実現するためにいかなる種類の与件データが必要なのかを知ること，そして(2)与件データと意思決定後の結果との間にいかなる因果関係が存在するのかを予測すること，さらには(3)意思決定の質の向上を低コストで可能とするデータ活用の仕組みを設計することである．

これら3つの課題はそれぞれがお互いに強く影響し合う関係にあり，個別的に課題を解決することが困難である．つまりスコアリング技術を産業の経営インフラとして有効活用するには，優れてシステマティックな戦略を構築することが必要となるのである．以降では様々なスコアリング法を例に，この課題の解決法について説明させていただく．

しらかわ　ひろし
東京工業大学 理財工学研究センター
〒152-8552 東京都目黒区大岡山2-12-1

2. クレジットスコアリング

クレジットスコアリングの起源は，1956年に創業されたFair Isaac社によるクレジット業界向けの与信意思決定支援サービスに始まる．ちなみに世界ではじめてのクレジット会社であるDiners Clubが創業されたのが1950年であり，わずか6年前にはじまった新産業においてスコアリングによる与信意思決定の向上がクレジット産業における成功の鍵であることを予見したFair-Isaac両氏の慧眼には恐れ入る．

このFair Isaac社の提供するスコアはおのおののデータベースごとの母集団の差を反映して個別名称が異なるものの，全体としてFICOスコアと総称されている．このスコアにもとづいてクレジット各社がその会員に対する与信判断（正確には，加盟店のクレジット会員に対する売掛金のノンリコース型ファクタリング）を行うことにより，個人顧客に対する定型的意思決定の結果をその客観的な個別信用情報（クレジットレポートと呼ばれる）から統計的に予測するのである．

ちなみにクレジットレポートの情報からどのように与信リスクと因果関係のあるファクター群を抽出するのかは，長い間，Fair Isaac社からは公開されていなかった．しかし被与信者並びにデータベース会社からの強い要請により，昨年ようやく利用されているファクターの概要が公開された．米国で個人情報保護が理解され，公正信用報告法（Fair Credit Report Act）が制定されたのが1970年であるから，信用リスクについての情報公開の社会的仕組み創りが始まってから30年後に，ようやくデータの価値を引き出す方法に対しても情報公開が行われたことになる．

3. クレジットスコアリングから金融データマイニングへ

クレジットスコアリングの背景にある大量のデータから価値のある情報を引き出す技術は，今日ではデータマイニングと呼ばれている．データマイニングが経

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

営意思決定の技術として1995年以降着目されたのには，以下の3つの理由がある．第1に，IT革命や人工知能の急激な進展に伴い，情報収集や情報処理コストが格段に安くなり，大量のデータを集めかつそこから有益な情報を引き出す仕組みをリーゾナブルなコストで構築可能となったこと，第2に近年の経営環境の変動性の増大に伴い経験則の有効期間が急速に短命化し，逐次更新される最新データに応じた最良の定型的意思決定法がタイムリーに必要となったこと，そして第3に企業経営環境の競争激化に伴い，顧客ごとのデータに基づくセグメンテーションの徹底により可能な限りリスクを削減しリスク1単位当たりの資本収益性を向上する必要にせまられたことである．

この結果現在の多くの金融機関においては，勘定系と呼ばれる決済システムをサポートする基盤情報システムに加え，情報系と呼ばれる金融データマイニングによる意思決定の質の向上をもたらすアプリケーション構築の重要性が高まりつつある．このようなデータマイニング技術の有効活用に最も重要となる原材料が，データウエアハウスと呼ばれる組織化されたデータベースである．

しかしその一方，経営意思決定の質の向上をもたらすデータセットとしてどのようなものが必要なのかは，データマイニングを施した結果，徐々に明らかになるものである．また従来はあまり有効でなかったデータ項目が，いつ有効なシグナルをもたらすデータとなるのかも予想はつきにくい．このためデータベース構築の初期時点では，コントロールしたい結果と因果関係のありそうな入力データセットをできるだけ大きく想定し，その中から意思決定の質の向上をもたらすキーファクターをデータマイニング作業により徐々に抽出する必要がある．この意味でデータウエアハウス構築とデータマイニングは密接に連携し，その熟成は相互補完的に行う必要があるのである．

4. 中小企業金融へのスコアリング技術の適用

近年，銀行等の間接金融部門が，直接金融の比重を高めつつある大企業向けの与信市場から，ミドルリスクミドルリターンと呼ばれる中小企業向け与信市場へと融資ビジネスの軸足を移動しつつある．

従来中小企業向け与信は与信額当たりの審査コストが高いこと，中小企業側からの情報公開が不充分でありリスク審査が困難であることなどにより，その潜在的成長性は認識されていたものの低コストで効率的に優良な与信案件を発掘することができなかった．このような状況において，金融ビックバンやIT及び人工知能技術の進展に伴いスコアリング技術を利用することによって中小企業金融を活発化させようという動きが始まりつつある．

しかし金融データマイニングの技術を中小企業金融等に利用するには，その本来の利用目的から生じる本質的な問題があることは，専門家の間では良く知られている．これは与信意思決定の質を向上させるために必要な学習データとして，本来は回避すべきグレーもしくは不良な与信案件の学習データを，安定的な予測に十分なサンプル数だけ必要とすることである．この問題は各金融機関ごとに与信ビジネス競争をしても解決できず，個々の銀行とは独立な横断的な不良与信案件データ共有の仕組みを必要とすることは明らかである．

この動きに先鞭をつけたのが，先述のFair Isaacs社により展開された中小企業向けスコアリングシステムSBSS（Small Business Scoring Service）である．SBSSでは中小企業向けの与信意思決定の質を向上させるために，全米商業銀行のリスク管理活動を推進するRMA（Risk Management Association）と協力し，大量の不良与信データをもとに中小企業向けデータマイニングシステムを構築した．この与信エンジンは当初は驚くべき与信効率を生み出し，近年Wells Fargo銀行が飛躍するきっかけを作ったといわれている．

その後経済リセッション状況を迎え景気変動に対するシステマティックリスクによる与信判断の脆弱性も指摘されており，それが長期的な与信意思決定の質の向上を本当に実現していたのか否か，今後の動きを慎重に見極める必要があろう．しかし不良与信案件のデータ共有とデータマイニング技術の適用により，従来と比較して短期中小企業与信の意思決定の質が向上できたことはほぼ間違いなさそうである．

5. CRAFTスコアリング法

理財工学研究センター（CRAFT, Center for Research in Advanced Financial Technology）では，㈱東芝研究開発センター，㈱和陽インターナショナルコンサルティング，太陽監査法人等と協力して，CRAFTスコアリング法と呼ばれる企業財務リスクの標準的評価技術を開発・公開した．CRAFTスコアリング法は，産業間金融の担い手に対する与信意思決定

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

の質の向上を目指しており，標準的な財務分析において用いられている企業診断カルテ項目のうち，財務情報並びに質的ファクト情報を中心に76項目の指標を取り上げ，指標測定後の一年以内の倒産事象との因果関係をシステマティックに分析する．

CRAFTスコアリング法では，各種財務データの統計的な説明力を裏づけとすることにより，与信者・被与信者の両者から中立的で透明性のあるわかりやすい評価が可能となる．また評価モデル構築後は，取引先の連続2期間の財務諸表を入手できれば，手軽に企業格付けおよび倒産確率が評価できる．以降では，CRAFTスコアリング法の具体的な計算手続きを，概説する．

【ステップ1】

取引先の直近財務データ $\boldsymbol{X}_t=(X_{1t}, X_{2t}, \cdots)$ と引き続く1年以内の倒産の有無を示すデータ $Y_t(=0$；非倒産，$=1$；倒産）の時系列データセットに基づき，モデル推定用の学習セットと，モデル評価用のテストデータセットを作る．倒産データが少ない場合には，倒産データに適切にブースティング処理を行う．

【ステップ2】

ロジット型の倒産確率モデル

$$P[Y_t=1|\boldsymbol{X}_t=\boldsymbol{x}]=\frac{1}{1+\exp\{-(\alpha+\beta^T\boldsymbol{x})\}}$$

により，学習データに適合する確率モデルを，各モデルパラメータ数ごとに，Forward変数選択基準に基づいて最尤法（平均対数尤度最大化）により推定する．

【ステップ3】

推定モデルの学習データに対する平均対数尤度とテストデータに対する平均対数尤度を比較し，両者のモデルパラメータ数に応じた差を，線形モデルにより評価する．

【ステップ4】

ステップ3で得られた偏差を考慮して，学習データについての自由度調節ずみの平均対数尤度を計算し，その結果に基づいて適切なモデルを選択する．

【ステップ5】

予想平均倒産確率に応じて倒産データの重みを変更した学習データに対し，ステップ4で選択されたモデルのスコア値 $\alpha^*+\beta^{*T}x$ の分布を求める．

【ステップ6】

クラスター数をパラメータとして，以下のステップ6-1から6-3を繰り返し，ステップ6-3で求められる2乗誤差を最小化する格数（＝クラスター数＋1）を，適切なモデル格数として採用する．

【ステップ6-1】

ステップ5で得られたスコア値分布に対して級内分散最小化基準に従ってクラスタリングを行い，スコア値のノンパラメトリック分布を離散近似する

【ステップ6-2】

ステップ5で得られた学習データについて，ステップ6-1で得られたクラスタリングをもとに，期首・期末間のクラスタ変動を計測し，倒産状態を含めた格付け状態遷移確率行列を推定する．さらに格付け状態ごとの融資額1円当たりの期待割引回収価値を1とする適性金利を評価する．

【ステップ6-3】

予想平均倒産確率に応じて倒産データの重みを変更したテストデータに対し，ステップ6-1のクラスタリング結果に対応した格付けを行い，さらにステップ6-2で計算された適正金利に基づいて融資額1円当たりの割引回収価値を格付け状態ごとに計算し，理論上の期待値1との2乗誤差の加重総和を求める．

以上の手続きからお分かり頂けるように，CRAFTスコアリング法の特徴は，学習データのみの推定によるオーバーフィッティングをテストデータによる評価で克服し，予測精度を考慮した格付けモデル選択を可能とした点にある．

以降では，㈱東京商工リサーチの「TSR DATA BANK SERVICE MT & FD 2000年版」に基づいて，1996年から2000年の間の5連続財務データを持つ建設業，製造業，および卸・小売業・飲食店計約20,000社（うち3,427件は分析期間中の倒産企業）に対し，CRAFTスコアリング法による倒産確率の推定を行った結果を紹介する．

まず【ステップ1】のデータ分割において，学習データ用に1997年3月末の直近の指標情報及び1997年4月～1998年3月末までの倒産・非倒産属性，テストデータ用に1998年3月末の指標情報及び1998年4月～1999年3月末までの倒産・非倒産属性を抽出する．

ついで【ステップ2】において，学習データに対する平均尤度を最大化する指標集合をForward基準により選択し，モデルパラメータ数に応じて最大化した対数尤度の計算結果を，図1に示す．上側の線は，学習データに対する最大平均対数尤度を表し，下側の線は同じ推定値のもとでのテストデータの平均対数尤度を表している．両者の間には，モデルパラメータ数に応じてギャップがあることがわかる．

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

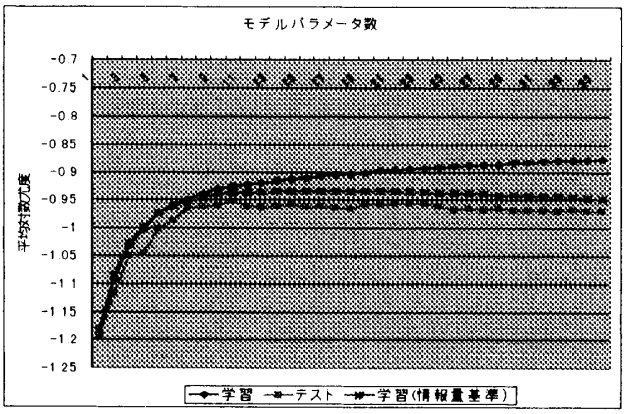

図1　モデルパラメータ数に応じた平均対数尤度

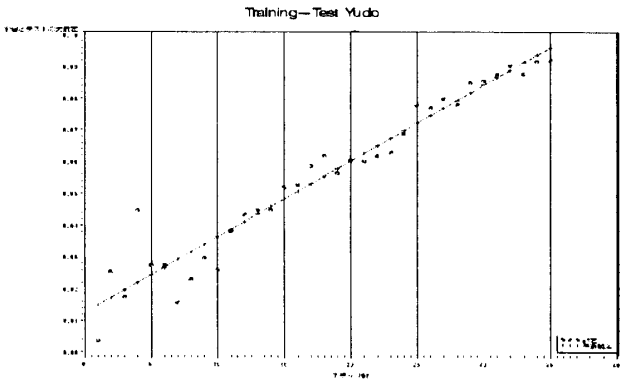

図2　モデルパラメータ数に応じた学習データとテストデータに対する平均対数尤度の差

【ステップ3】に従って，このギャップを示したのが，図2である．ギャップ値がモデルパラメータ数に関する線形関数でうまく近似できることが，データから読み取れる．この推定された偏差をもとに，学習データに対する最大平均対数尤度関数を修正したのが，図1の中側の線である．通常モデル選択の基準として用いられる情報量基準（AIC）と比べ，図2で計算上求められる偏差修正項は約60倍の大きさを持つ．これは，真の倒産確率分布が当初考えたロジット型分布族とは異なるクラスであることを示唆している．

さらに【ステップ4】により，自由度調節済みの最大平均対数尤度を大域的に最大化するモデル選択を考えると，パラメータサイズ＝18が選択され，その具体的な指標集合は，表1のようになる．結果として，倒産事象との間に重要な因果関係が存在すると考えられる指標が，自動的に抽出されていることがわかる．

ステップ4で確定した最適な指標集合に対し，【ステップ5】で，学習データのスコア値の度数分布を求めた結果を図3に示す．上側の大きな山が非倒産企業に対する度数分布，下側の小さな山が倒産企業に対す

表1　最適なモデルの指標集合

NO.	指　　標
1	外部負債依存率
2	有利子負債平均金利
3	買入債務回転期間
4	一人当り売上高
5	総資産運転資本率
6	一人当り経常利益
7	当座比率
8	売掛金回転期間
9	受取手形（割引前・譲渡前）回転期間
10	資本回転期間
11	売上高営業費用率
12	社債・長期借入金回転期間
13	有形固定資産回転期間
14	総資本増加率
15	一人当り売上総利益
16	棚卸資産回転期間
17	経常利益増加率
18	固定負債回転期間

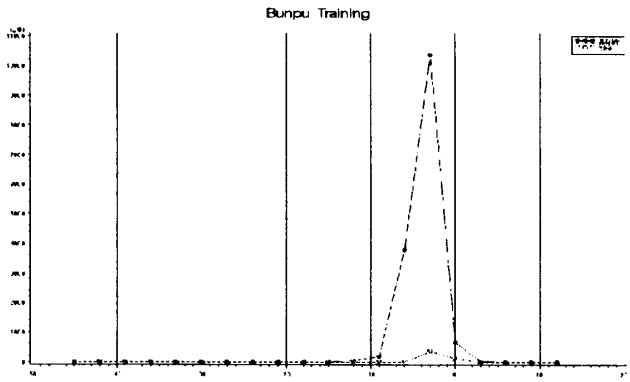

図3　学習データのスコア値度数分布

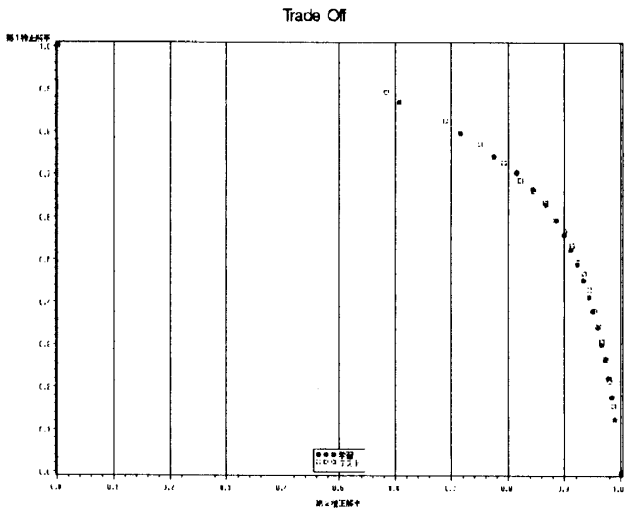

図4　第1種正解率と第2種正解率のトレードオフ

る度数分布である．またこのスコア値が，引き続く1年以内の倒産事象とどの程度の因果関係があるのか，予測精度の観点から分析した結果を，図4に示す．X軸は第1種正解率（非倒産企業を非倒産と予測），Y

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

軸は第2種正解率（倒産企業を倒産と予測）を表す．このとき当該企業のスコア値<カットオフ値であれば非倒産と予測し，逆にスコア値>カットオフ値であれば倒産と予測した場合の，（第1種正解率，第2種正解率）の軌跡を，スコアのカットオフ値を徐々に変化させてプロットした軌跡をしめす．結果として，学習データ（●表示）のみならず，テストデータ（□表示）についても平均80%程度の予測精度を達成できることがわかる

引き続き【ステップ6】では，様々なクラスター数に対し，ステップ6-1からステップ6-3を繰り返し計算する．以降では特にクラスター数=9の場合について説明する．まず【ステップ6-1】で，この2つの分布を予想平均倒産確率のもとで重ね合わせ，級内分散最小化基準によりクラスタリングを行った結果が図5である．またこのクラスタリングにより計算された，各クラスのセントロイド及びレンジを表2に示す．この結果，各企業のある時点での財務指標を元にスコア値を計算し，クラスタリング結果への対応付けを行えば，当該時点での企業格付けが可能となる．

このようにしてステップ5で得られた学習データの期首・期末のスコア値について，【ステップ6-2】に基づき，所与のクラスター数に倒産状態1を加えた格数=10のもとで，格付け状態遷移確率行列を推定した結果が，表3である．この格付け状態遷移確率行列から，格付け状態が悪くなる（=スコア値が大きくなる）ほど，倒産確率が高くなる傾向があることが，読み取れる．さらにこの遷移確率行列から，安全資産の金利=5%，倒産時の元本回収率=0%と仮定したもとでの，任意期間の，融資額1円当たりの期待割引回収価値が，当初の投資額である1円と一致する適正金利を各格付け状態ごとに求めた結果を，図6に示す．最上位の曲線が，第9格付けに対する適正金利の期間構

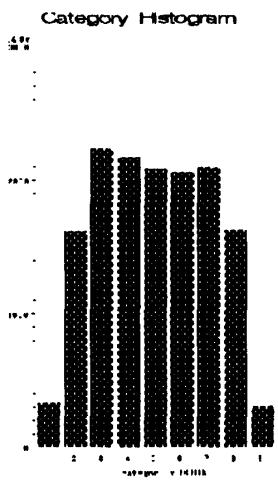

図5　クラスタリング後の離散度数分布

表2　格付け状態（クラスタ）ごとのスコア値のレンジ

格付け状態	セントロイド	クラスターのレンジ
1	-7.114429311	-5.639435186
2	-4.423269274	-3.804315466
3	-3.340536489	-2.955953926
4	-2.667445056	-2.393279139
5	-2.157838905	-1.927192291
6	-1.68777242	-1.442639437
7	-1.150417872	-0.821994051
8	-0.413749223	0.258871785
9	1.086619083	

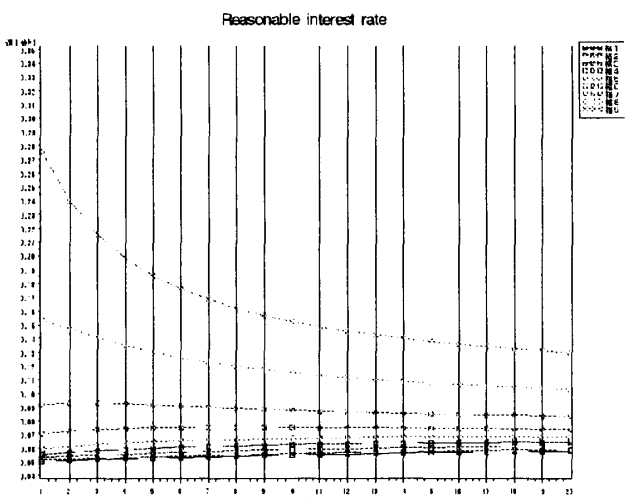

図6　格付けごとの適正金利の期間構造

表3　格付け状態遷移確率行列の推定結果

格付け	1	2	3	4	5	6	7	8	9	10
1	0.726994	0.196319	0.046012	0.006135	0.01227	0.003067	0.003067	0.003067	0.003067	0
2	0.050095	0.689918	0.209892	0.033608	0.008244	0.004439	0.001268	0.000634	0.000634	0.001268
3	0.00357	0.168228	0.552432	0.195002	0.051316	0.015172	0.006693	0.00357	0.000892	0.003124
4	0.001389	0.023148	0.212963	0.461111	0.205093	0.065278	0.018056	0.006481	0.001389	0.005093
5	0.000491	0.011291	0.062347	0.233186	0.418753	0.201767	0.049092	0.012764	0.001473	0.008837
6	0.000469	0.002813	0.019691	0.060947	0.23488	0.42616	0.200656	0.033286	0.001875	0.019222
7	0	0.004076	0.005888	0.028533	0.052536	0.212409	0.505435	0.141757	0.011322	0.038043
8	0	0.001112	0.003337	0.005006	0.017798	0.037264	0.208009	0.579533	0.058398	0.089544
9	0	0	0.004717	0.004717	0.007075	0.033019	0.049528	0.349057	0.375	0.176887
10	0	0	0	0	0	0	0	0	0	1

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

造を表し，最下位の曲線が，第1格付けに対する適正金利の期間構造を表している．第8格付け及び第9格付けの期間構造が高いものの，第7格付け以下の期間構造は，約10%以下（リスクプレミアムが5%以下）の水準で推移している．

ついで【ステップ6-3】に従って，テストデータに対し，仮に2年間の融資について，適正金利のもとでの融資額1円当たりの割引回収価値を格付け状態ごとに計算し，1円との2乗誤差の予想平均倒産確率に応じて加重した総和を求めた結果が，図7の格数＝10（クラスター数＝9）に対応する値である．さらにステップ6-1から6-3をクラスター数1から15まで繰り返し計算した結果が，図7の変化に対応する．これから，2乗誤差の最小値を実現するクラスター数は14であるものの，結果の安定性を考えると，揺らぎが起こる前に達成されている最小値近辺の格数を選択することが望ましいことがわかる．

以上の手続きにより，最適な指標集合，スコア値関数，及びスコア値の最適な格付け数へのクラスタリング，最適な格付け状態遷移確率行列等，企業信用リスクの評価に必要なモデル構造が確定する．

CRAFTスコアリング法は，2001年度の情報処理振興協会の中小企業経営効率改善支援ソフトウエア開発・実証事業において採用され，そこで開発された与信リスク管理支援ツールには，上述の計算結果に基づいた自動格付けシステムが組み込まれている．この与信リスク管理支援ツールは，理財工学研究センターのホームページhttp://www.craft.titech.ac.jpから無償でダウンロードすることができ，2001年7月末現在，日本の600社以上の中小企業において活用されている．

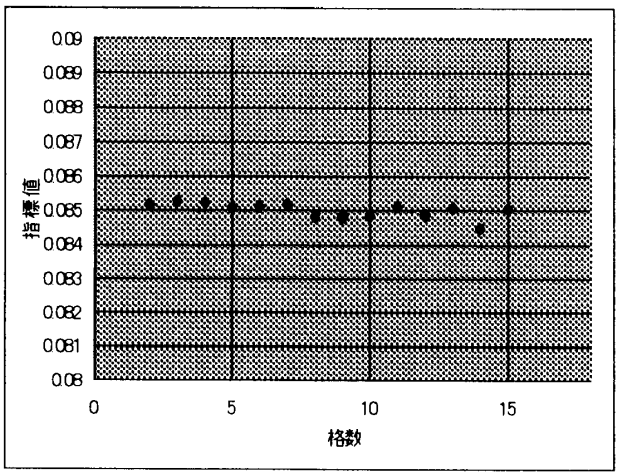

図7　テストデータの割引回収価値の2乗誤差

6. スコアリングとe-コマース

分散経済システムにおける経済成長の鍵は，どのような経済主体と，どの程度の分業を行うかというジョブディスパッチングの動的な選択にある．このジョブディスパッチングは，結局のところある企業がどのような与信先を選ぶかということの積み重ねにより決定する．したがってスコアリング技術により，与信意思決定の質を地道に向上することが，実体経済のサービス品質を改善し，経済成長のポテンシャルを引き出す上で非常に重要となるのである．

近年，このようなスコアリング技術のメリットを生かす上で重要な，経済システムの変革が起きつつある．それはe-コマースによるインターネット上での企業間連携の構築である．e-コマースでは，従来の商取引では困難であった3つの機能を容易に実現することが可能である．第1に，様々な個別的ニーズに応じた取引機会を広範囲に集めることにより，迅速に所望の取引を実現する流動性の創造機能，第2に，すべての取引機会の中からもっとも有利な提案を探索する効率性の創造機能，そして第3に，これらの取引の信頼性を評価するための既往の全取引を含めた情報共有機能である．

これら3つの機能は，実はオフラインの世界で地道に行われている共有データウエアハウスの構築と新規与信時におけるスコアリング分析を，e-取引の場面で広範囲かつリアルタイムに実現することに他ならない．しかし，このような企業間情報共有のためのソフト・インフラが構築されれば，未知の企業間連携の効率化が隅々まで実現され，結果として資本利用効率のよい企業同士の連携が加速されることで，従来の経済システムでは実現不能な，非常に高い経済成長率が達成可能となるのである．

このようなインターネット上での企業間連携を促進するのに，非常に有望と考えられている技術が，XMLによる企業財務情報の共有化である．特に，米国の公認会計士協会が中心となって推進しているXBRLと呼ばれる，財務諸表等の情報共有のための共通DTDを構築しようとする試みは，現在Ver 2.0まで進んでおり，近い将来，企業財務会計報告における標準的なe-レポーティング手段になると考えられている．

例えばCRAFTスコアリング法で紹介したスコア値は，外部会計報告の都度変更されるに過ぎない（多

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

くの中小企業の場合，1年に1回)．しかしe-レポーティング機能の実現により，月次ベースの外部会計報告が可能となれば，スコア値も月次ベースで改定可能となり，より実態にマッチした与信意思決定のコントロールが可能となる．この意味でスコアリング技術は，e-経済システムが実現した環境下において，もっともその効果を発揮するといえよう．

謝辞　本原稿をまとめるに当たって，㈱東芝研究開発センターの西川武一郎研究員との共同研究成果を一部引用させていただきました．末筆ながら，深く感謝の意を表します．

参考文献

[1] Altman, Edward. (1968), Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, Journal of Finance, 23(4), 589-609.

[2] 上坂吉則，尾関和彦 (1990)，パターン認識と学習のアルゴリズム，文一総合出版．

[3] 木島正明，小守林克哉 (1999)，信用リスク評価の数理モデル，朝倉書店．

[4] 白田佳子 (1998)，企業倒産予知情報の形成に関する研究，東洋経済新聞社．

[5] 白川浩，西川武一郎，梅木直子 (2001)，CRAFTスコアリング法，CRAFT Technical Report Series.

[5] 豊田秀樹 (2001)，金鉱を掘り当てる統計学－データマイニング入門－，講談社．

[6] ジョセフ・P・ビーガス (1997)，ニューラル・ネットワークによるデータマイニング，日経BP社．

[7] エリザベス・メイズ編 (2001)，クレジット・スコアリング，シグマベイズキャピタル．

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．