日本オペレーションズ・リサーチ学会
2004年秋季研究発表会

多目的遺伝的アルゴリズムを用いた財務スコアリングモデルのチューニング

03000230 MTEC／東京工業大学・理財工学研究センター 山内 浩嗣 YAMAUCHI Hiroaki

1. 概要

本研究では財務スコアリングモデルのチューニング方法として、遺伝的アルゴリズム(GA)を応用した方法を提案する。

GA による解探索においては多目的 GA の応用を試み、信用リスク計量モデルを評価する複数の評価尺度を目的関数に設定した。一方、頑健性の高い解を生成するための工夫として(サンプルに対して出来るだけモデルがオーバーフィットしないための工夫として)、与えられたサンプルの格付構成比を保ちつつ復元抽出するリサンプリングを世代ごとに複数回実行し、それらリサンプリングデータを入力データとした場合の目的関数値の平均で解の適応度を評価した。そして頑健性の高い解を選択するための情報を得るために、最終世代の解に対しては目的関数値のブートストラップ統計量を計算した。

スコアリングモデルとはスコアリングテーブル(配点表)と呼ばれる変換テーブルに従って財務指標を点数化し、複数の指標の点数を合計して信用リスク量の評価値とするモデルのことである。この手法は銀行における与信先企業の財務力評価などの用途で従来から広く用いられている。

指標の区切り	0%未満	0%～30%	30%以上
スコア値	-30 点	0 点	20 点

≪ある財務指標のスコアリングテーブルの例≫

本研究で提案するチューニング手法は、一般的な信用リスク計量モデル構築プロセスの**手順 4** と**手順 5** に該当する。

≪信用リスク計量モデルの一般的な構築プロセス≫

手順 1　モデルの選択
手順 2　選択したモデルに適した分析用データの準備
手順 3　財務指標の選択
手順 4　モデル・パラメターの決定
手順 5　バックテスト等によるモデルの頑健性検証

なおスコアリングモデルにおける**手順 4** とはスコアリングテーブルのチューニングを意味し、具体的には以下のことを行う。

手順 4-A　各指標に対する配点の決定
手順 4-B　スコアリングテーブルの区切値とスコア値の決定

2. モデル評価関数の設定

信用リスク計量モデルには多くの評価尺度がこれまでに考案されてきた。本研究では評価関数(目的関数)として、実務的観点から正判別率 CDR と序列付け関数 OF の 2 つを採り上げた。なお OF は今回新たに考案したものである。

$$CDR(\mathbf{s}) \equiv \frac{\displaystyle\sum_{i\in\text{正常企業}} I\{x_i(\mathbf{s}) \ge CV(\mathbf{s})\} + \sum_{i\in\text{デフォルト企業}} I\{x_i(\mathbf{s}) < CV(\mathbf{s})\}}{N}$$
$$= 1 - \text{誤判別率}$$

$$OF(\mathbf{s}) \equiv \frac{\displaystyle\sum_{j=1}^{RMAX-1} (N_j + N_{j+1}) \cdot \left|\bar{x}_j(\mathbf{s}) - \bar{x}_{j+1}(\mathbf{s})\right| \cdot f\left(\bar{x}_j(\mathbf{s}) - \bar{x}_{j+1}(\mathbf{s})\right)}{\displaystyle\sum_{j=1}^{RMAX} \sum_{i\in\text{格付}j\text{の企業}} \left|x_i(\mathbf{s}) - \bar{x}_j(\mathbf{s})\right|}$$

ただし、

スコアリング関数: $S(\mathbf{s}, v_i^p) = s_1^s(p) \cdot I\{v_i^p < s_1^b\} + \sum_{k=1}^{nb-1} s_k^s(p) \cdot I\{s_k^b \le v_i^p < s_{k+1}^b\} + s_{nb+1}^s(p) \cdot I\{s_{nb}^b \le v_i^p\}$

企業iの合計スコア: $x_i(\mathbf{s}) = \sum_{p=1}^{nv} S(\mathbf{s}, v_i^p)$

正常企業の平均スコア: $\bar{x}_n(\mathbf{s}) = \sum_{j=1}^{RLB} N_j \bar{x}_j(\mathbf{s}) \Big/ \sum_{j=1}^{RLB} N_j$

デフォルト企業の平均スコア: $\bar{x}_d(\mathbf{s}) = \sum_{j=RLB+1}^{RMAX} N_j \bar{x}_j(\mathbf{s}) \Big/ \sum_{j=RLB+1}^{RMAX} N_j$

判別閾値: $CV(\mathbf{s}) = \left(\bar{x}_n(\mathbf{s}) + \bar{x}_d(\mathbf{s})\right)/2$

格付内平均スコア: $\bar{x}_j(\mathbf{s}) = \frac{1}{N_j} \sum_{i\in\text{格付}j\text{の企業}} x_i(\mathbf{s})$

$$I(z) = \begin{cases} 1 & \cdots z\text{が真のとき} \\ 0 & \cdots z\text{が偽のとき} \end{cases}$$

$$f(z) = \begin{cases} 1 & \cdots z \ge 0 \\ P & \cdots z < 0 \end{cases}$$

v_i^p = **企業iの財務指標pの値**
N_j = **格付jに属する企業数**
N = **全企業数** = $\sum_{j=1}^{RMAX} N_j$

R_i = **企業iの格付数値**
RLB = **正常企業の格付数値の上限**
$RMAX$ = **サンプル全体の格付数値の最大値**
P = **格付平均スコアの序列が逆転していた場合のペナルティ**(>0)

なお格付数値とは格付記号AAA, AA+,…を1,2,…という自然数に置き換えたものであり、**s** はスコアリングテーブルを表す以下のベクトルである。

$$\mathbf{s} \equiv \Big[\underbrace{\underbrace{s_1^b(1),\cdots,s_{nb}^b(1)}_{\text{指標1の区切り値}},\underbrace{s_1^s(1),\cdots,s_{nb+1}^s(1)}_{\text{指標1のスコア値}}}_{\text{財務指標1のスコアリングテーブル}}, \quad \cdots \quad ,s_{nb+1}^s(nv)\Big]^T$$

3．多目的最適化問題としての定式化

スコアリングモデルのチューニングを定式化する。

$$\underset{\mathbf{s}}{\textbf{Maximize}} \quad [CDR(\mathbf{s}), OF(\mathbf{s})]$$

Subject to 区切り値の下限値 $\leq s_k^b(p) \leq$ 区切り値の上限値

$$for \quad k = 1,2,\cdots,nb \quad p = 1,2,\cdots,nv$$

スコア値の下限値 $\leq s_j^s(p) \leq$ スコア値の上限値

$$for \quad j = 1,2,\cdots,nb+1 \quad p = 1,2,\cdots,nv$$

$$s_{m(i)}^i(p) < s_{m(i)+1}^i(p)$$
$$for \quad i = b,s \quad p = 1,2,\cdots,nv$$
$$m(b) = 1,2,\cdots,nb-1 \quad m(s) = 1,2,\cdots,nb$$

$$s_{m(i)}^i(p) \in \mathbf{Z}$$
$$for \quad i = b,s \quad p = 1,2,\cdots,nv$$
$$m(b) = 1,2,\cdots,nb \quad m(s) = 1,2,\cdots,nb+1$$

nvは財務指標数、nbは個々の財務指標に対する区切り値の数である。今回はnb=10、区切り値の下限=-24、上限=+24、スコアの下限=-20、上限=+20と設定した。なお区切り値が-24から+24までの整数値をとることに対応するため、各財務指標を昇順ソートしてから銘柄数で51等分し、区分を割り当てた。

4．GA 適用のポイント

個体数 各世代の個体数は500とした。

解の定義 本研究では先に定義したスコアリングテーブルを表すベクトル$\mathbf{s}$を解(染色体)として定義した[1]。

適応度の定義 本研究ではFonsecaらによる定義を用いた。

個体xの適応度 ≡ 個体数 − 個体xのパレートランク

リサンプリング 世代交代ごとに、与えられたサンプルの格付構成比率を保ちつつ復元抽出するリサンプリングを複数回行い、データセットを複数(今回は 2 セット)用意した。これらに対して解ごとに目的関数値の平均を求め、適応度を評価した。

複製選択 エリート保存戦略＋重複ありのルーレット戦略(＋線形スケーリング法)によって交叉する親を選択した。

交叉 本研究では一様交叉を採用した[2]。生成される子に偏りが生じないよう、交叉時にはベルヌーイ試行でCDRかOFを毎回選択し、選ばれた評価値で親をソートしてから交叉させた。

突然変異 すべての遺伝子座に対して「突然変異確率=1/遺伝子座数」で突然変異を発生させ、区切り値あるいはスコア値の上下限内で一様乱数を生成して遺伝子の値を置き換えた。

生存選択 親＋子の解集合の中でパレートランク1の個体数が500未満ならそれらを無条件に選択し、残りはパレートランク2以上の個体だけによるルーレット戦略(重複なし)で選択した。パレートランク1の個体数が500を超えた場合は、クーロンポテンシャルの考え方を応用したシェアリングを行った。

終了判定 本研究では特に収束条件を設定せず、必ず700回の世代交代を行った。

最終世代の解評価 頑健性の高いモデルを選択するための基準として、モンテカルロ法による1000回のブートストラップを行い、目的関数値の平均・標準偏差・バイアスを求めた。

5．データの概要

財務指標データは金融業を除く東証一部企業の 2001 年 2 月～2003 年 8 月決算の単体データを用いた。数 10 種類の財務指標候補から、統計モデルによる指標選択や定性的判断などにより以下の 7 指標を選択した。従ってnv=7 である。格付は R&I が公表しているものを用いた。今回は格付取得企業をサンプルとしたため、実際にデフォルトした企業はごく僅かであった(延べ 2023 社中 7 社)。従ってここでは擬似的にデフォルトを定義し、AAA 格～BBB-格を正常企業(延べ 1705 社)、BB+格以下をデフォルト企業(延べ 318 社)と見なした。

対数売上高　自己資本比率　固定長期適合率　当座比率
総資本経常利益率　売上高当期利益率　総資本回転率

6．分析結果

分析結果については当日発表します。

7．参考文献

[1] 三宮信夫・喜多一・玉置久・岩本貴司、「遺伝アルゴリズムと最適化」、朝倉出版、1998 年.

[2] 山内浩嗣、「多目的遺伝的アルゴリズムを用いたスコアリングモデルのチューニング」、日本金融証券計量工学学会(JAFEE)第 21 回大会予稿集、2004 年 8 月.

≪発表者の連絡先≫
yamauchi@mtec-institute.co.jp（MTEC）
yamauchi@craft.titech.ac.jp（東工大：2004 年 11 月末まで）

---

1 制約条件より、すべての遺伝子の取りうる値は整数値に限られるため、本問題は整数型 GA となる。

2 当初は局所探索を、親は交叉の前に、子は生存選択の前に行うことを想定していた。しかしリサンプリングによって入力データが世代ごとに置き換わるため、実際に分析してみると局所探索が有効に機能しないことが分かった。よって現時点では計算時間短縮のため遺伝的局所探索を行っていない。なお本研究での局所探索とはすべての区切り値・スコア値を 1 ノッチずつ上下にずらすこととし、移動戦略は最適移動戦略を採用していた。