進化論的手法を用いた金融データの予測

伊庭斉志

蒲田　涼馬 (Ryoma Gamada)
u020010@st.pu-toyama.ac.jp

富山県立大学　情報システム工学科

May 10, 2023

背景と目的

背景

遺伝的プログラミングというものがある.
　　　しかし, 従来の遺伝的プログラミングにはいくつかの問題が指摘されている.

目的

問題として挙げられている遺伝的オペレータを的確に制御する手法の欠如, ノードの設計時における表現の問題, 木構造の評価についてGP に GMDH を組み込むことでそれを改善.
金融データでそれの有用性をたしかめる.

従来の GP

GP(Gebetic Programming) 遺伝的プログラミング

GP は交叉や突然変異などの遺伝的なオペレータを気構造に適用し，目的とするプログラムや概念木を探索する手法.

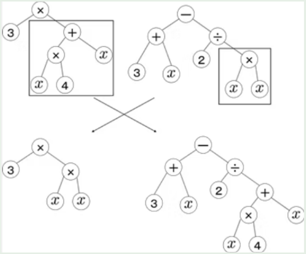


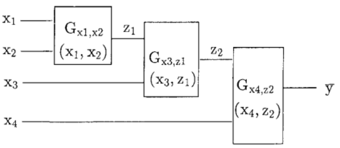
図 1: 実行結果

GP の問題点

1. 遺伝的オペレータを的確に制御する手法の欠如
2. ノードの設計時における表現の問題
3. 木構造の評価

GMDH

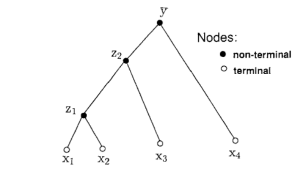
GMDH は, システム同定問題の解法に用いられる多変量解析手法であり, 出力関数 y を近似するためのネットワークを構成する.
1. 入力変数を $x_1, x_2, ..., x_m$, 出力変数を y とする.
$VAR := x_1, x_2, ..., x_m$ を初期化する.
2. VAR から 2 つの要素 z_a, z_2 を取り出す. 他重回帰分析を用いて出力 y を z_1, z_2 により最小誤差で近似する式 $G(z_1, z_2)$ を構成する.
3. z の近時の制度が, 与えられた基準以上のとき y の完全系を z として終了する.
4. でなければ $VAR := VAR \vee \{z\}$ として 2 へもどる



(a) GMDH ネットワーク

木のノードの係数を再計算

GMDH で与えられる完全形 $\bar{y}$ は木の形式で表現される.
この場合完全形は $\bar{y} = G_{x_4, z_2}(x_4, z_2)$
$z_1 = G_{x1,x2}(x_1, x_2)$, $z_2 = G_{x3,z1}(x_3, z_1)$
z_1, z_2 は GMDH アルゴリズムで得られる部分表現式であり, これらの部分表現式を非終端記号, 入力変数を終端記号と考えることで図のような木構造を得ることができる. これから z_1, z_2, y を抜き取り, 行列計算を行うことでパラメータを計算する.
これをオペレータを適用するたびに行う.



(b) GMDH の 2 分木表現

図 2: 例

MDL に基づく適合度計算

MDL は木の複雑さと近似の誤差のトレードオフを考慮した評価である.

木構造に対する MDL 値は

$MDL = 0.5N \log S_N^2 + 0.5k \log N$

$S_N^2 = \frac{1}{N} \sum_{i=1}^{N} |\bar{y}_i - y_i|$

MDL 値は最小値を最適値としてとるため, スケーリングウィンドウ手法で適合度に変換する.

スケーリングウィンドウ手法は個体の適合度をその個体と何世代か前の最悪個体の関数値の差として計算する方法である.

手法

今回の実験では GP と STROGANOFF, ニューラルネットワークで実行.
1993 年のデータを用いる.
1 分毎のデータを 0 から 1 の間をとるように正規化する
$sdata = \frac{data - min}{max - min}$
$data$ は正規化前のデータ, $sdata$ は正規化後のデータ
さらに正規化されたデータを訓練データ (最初の 3000 分) とテストデータ (それ以降) とに分ける.
木構造を用いて予測値 $Pr(t)$ を出力, その値次第で売買を決定.
これをテストデータの数だけ行う.
A G の 7 通りの終端記号の元で実行した.

- 条件 A：{ y1, ... , y10, $\Re$ }
- 条件 B：{ ave1, ... , ave10, $\Re$ }
- 条件 C：{ m1, ... , m10, $\Re$ }
- 条件 D：{ ave1, ... , ave10, m1, ... , m10, $\Re$ }
- 条件 E：{ y1, ... , y10 ave1, ... , ave10, m1, ... , m10, $\Re$ }
- 条件 F：{ d1, ... , d10, $\Re$ }
- 条件 G：{ v1, ... , v10, r1, ... , r10, $\Re$ }

図 3: 終端記号

結果

GP の結果

条件	訓練データ内での平均二乗誤差	的中率 (%)		利益 (円)	
		平均値	最良値	平均値	最良値
A	1.79e-06	55.02	62.78	12411.01	31256.06
B	1.22e-05	47.47	48.17	-4093.22	-2341.50
C	5.82e-04	50.42	51.00	127.03	305.13
D	1.28e-05	41.09	51.64	-19727.52	-3811.19
E	1.94e-06	49.45	50.34	-806.33	1061.38
F	1.64e-06	55.02	62.49	14276.06	30722.81
G	1.80e-06	61.38	62.56	28942.03	30896.56

図 4: GP 結果

STROGANOFF の結果

条件	訓練データ内での平均二乗誤差	的中率 (%)		利益 (円)	
		平均値	最良値	平均値	最良値
A	9.37E-06	62.2	62.3	30516	30823
B	1.25E-05	57.3	57.7	18594	19194
C	6.57E-04	50.0	58.2	841	4044
D	1.25E-05	57.2	57.7	18390	19194
E	7.25E-04	51.2	51.3	5785	6071

図 5: STROGANOFF 結果

ニューラルネットワークの結果

隠れノードの数	平均 2 乗誤差 (訓練データ内)	的中率 (%)		利益 (円)	
		平均値	最良値	平均値	最良値
5	2.92e-06	58.2	60.2	23682	27586
10	2.70e-06	58.7	59.4	24725	26427
15	2.73e-06	58.3	59.5	23990	26245

図 6: ニューラルネットワーク結果

まとめ

まとめ

STROGANOFF は GP よりも利益を出すことができたのでこれを金融取引に用いるのは有用であると思われる.
ニューラルネットワークでも利益は出るが, 最良値は GP 系の手法より悪くなり, 局所値への陥りやすさが観察された.