

## クロスエントロピー最適化を用いた株価予測値の安定化手法

三浦和起<sup>†1</sup> 日野英逸<sup>†1</sup> 村田昇<sup>†1</sup>

時系列の予測は古くからある重要な問題であり、特に株価の予測は経済動向の予測や資産運用の指針として需要が高い。コンピュータ性能の発達と共に、学習理論を用いた経済時系列データに関する研究が活発に行われているが、株価のメカニズムを捉えることは依然として困難な問題である。本稿では、単一の予測モデルにより株価を一点で予測するのではなく、複数の予測モデルの学習を行い、各モデルに適切な重みを付けることで予測値の分散を低減する手法を提案する。基礎となる予測モデルは遺伝的プログラミングを用いて構成する。各予測モデルの重みは、学習用データと予測モデルの出力値とのクロスエントロピーが最小となるように定める。提案した予測手法の有用性を、人工データ及び日経平均株価の1分足の予測によって検証する。

### Stabilization of stock price prediction by cross entropy optimization

KAZUKI MIURA,<sup>†1</sup> HIDEITSU HINO<sup>†1</sup>  
and NOBORU MURATA<sup>†1</sup>

Prediction of time series data is a long standing important problem. Especially, prediction of stock price is much in demand for forecasting the economic trend and guideline for asset maintenance. Although there are growing number of studies on learning theory based time series prediction, the prediction of stock prices is still being very difficult task. In this study, the stock prices is predicted not only using one predictor, but using a set of predictors generated by the method of Genetic Programming (GP). Each element predictor is given non-negative weight, and the weight is optimized to minimize the cross entropy between the true learning stock prices and the weighted sum of predicted values. The proposed stock price prediction method is evaluated using both an artificial data and real-world stock price data.

<sup>†1</sup> 早稲田大学, Waseda University

## 1. はじめに

資金運用において株式投資は重要な選択肢であり、近年ではネット証券などの普及に後押しされ株式投資を行う人口が急増している。将来の株価を予測することは投資のリスクを回避するうえで非常に重要であり、伝統的な時系列解析からのアプローチ<sup>1)</sup>や、確率微分方程式に基づくアプローチ<sup>2)</sup>を代表として、多くの研究が行われている。一方、学習理論を用いた予測も盛んに研究されており、例えばニューラルネットワーク<sup>3)</sup>、サポートベクターマシン<sup>4)</sup>、遺伝的アルゴリズム<sup>5)</sup>、遺伝的プログラミング<sup>6)</sup>などを用いて株価予測モデルを学習する手法が提案されている。学習に基づくアプローチによって、ある程度実測値に近い予測値を得ることが可能であることが報告されている。

既存研究の多くは、単一の予測モデルから得られる単一の予測値の精度を向上させることを目的としている。一方、現実の株価等の経済時系列は様々な要因が絡み合い複雑な挙動を示すため、株価を単一の予測値で正確に予測することは一般的には困難である。そこで本研究では、複数の予測モデルを構築し、各々の予測モデルに対して適切な重みを設定することで、全予測モデルから得られる複数の予測値を用いて、分散の安定した分布を形成し、株価の挙動を捉える手法を提案する。予測値の分布が得られると、株価の予測値に対してその信頼性を自然に定めることが出来る。例えば多数の予測モデルの出力の平均値を株価予測値とする場合、出力の分散が小さい場合にはその予測は信頼性が高いと考えられ、実際の株式売買における指針の一つとなる。

学習に基づく予測モデルの構築は、通常は予測対象の株価(教師データ)と関連があると考えられる経済指標や株価の時系列データを入力データとして、教師データと予測モデルの出力値との誤差を最小化することで行われる。本研究では、複数の予測モデルの構築手法として、遺伝的プログラミング(Genetic Programming:GP<sup>7)</sup>)と呼ばれるメタヒューリスティクスを用いた。GPは、初期集団においてランダムに構築された個体(予測モデル:プログラム)が、生物の進化過程を模した自然淘汰、突然変異を含む進化(学習)を繰り返すことで、最終的に優れた予測モデルを出力する。また、最良の単一の予測モデルを求めるのではなく、複数の多様性を持った予測モデルを組み合わせることで予測値の分布を構築し、さらに各モデルに付与する重みの最適化により、適切なモデルを組み合わせる点が本研究の特徴である。

## 2. GPによる予測モデルの構築

本節では、予測モデルの生成に用いる Genetic Programming Polynomial Models<sup>8)</sup>を概

説する。GP の一般論及び本研究で利用するモデルの設定に関してはそれぞれ 7) と 8) に詳しい。

## 2.1 基本構造

GP は遺伝子型とプログラムを対応させ、進化を通して優れたプログラムを獲得する手法である。遺伝子型の表現としては木構造を用い、数式やプログラムのコードなどの構造を持ったデータを表現する。本研究では、図 1 に示すような各節が枝を 2 本持つ木構造を利用し、過去の  $\tau$  時刻分の株価  $\mathbf{x}_{t-1} = \{x_{t-\tau}, \dots, x_{t-1}\}$  を用いて、次の時刻の株価  $x_t$  を予測する。過去の  $\tau$  時刻分の株価からランダムに選ばれた幾つかの株価 (入力変数) を葉への入力とする。各ノードには伝達関数と呼ばれる 2 変数 2 次多項式  $y_i(u, v)$  が対応し、子ノード (あるいは葉) からの入力に対応する出力を親ノードへと渡す。表 1 に本研究で用いる伝達関数を示す。根に対応する伝達関数の出力値が、木構造モデルによる株価の予測値となる。この木構造で表現される関数が、GP における一つの個体となる予測モデル  $f_i(\mathbf{x}_{t-1})$  である。

各予測モデルにおいて、伝達関数の各係数  $a_j^{(i)}$  は、学習の段階で、ノードごとに実測値  $x_t$  と伝達関数の出力を用いて最小二乗法により定める。例えば  $\tau = 10$  時刻前までのデータ  $x_{t-10}, x_{t-9}, \dots, x_{t-1}$  を用いて  $x_t$  における時系列の値を予測する問題を考える。図 1 のノード  $y_6$  の例では、葉  $x_5, x_3$  には  $x_{t-5}, x_{t-3}$  が対応する。予測対象の時刻を  $t = T$  まで動かしたとすると、それに伴い  $x_{t-5}, x_{t-3}$  は、 $x_{T-5}, x_{T-3}$  までの値を得る。これらを各項で計算した値と実測値からなる正規方程式を解くことで各係数を定める。 $y_6$  の場合では、以下の様な方程式を解くことになる。

$$\begin{bmatrix} 1 & x_{t-5} & x_{t-5}x_{t-3} & x_{t-5}^2 \\ 1 & x_{t+1-5} & x_{t+1-5}x_{t+1-3} & x_{t+1-5}^2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{T-5} & x_{T-5}x_{T-3} & x_{T-5}^2 \end{bmatrix} \begin{bmatrix} a_0^{(6)} \\ a_1^{(6)} \\ \vdots \\ a_3^{(6)} \end{bmatrix} = \begin{bmatrix} x_t \\ x_{t+1} \\ \vdots \\ x_T \end{bmatrix} \quad (1)$$

このようにして各木 (モデル) の各ノードにおいて最小二乗法を用いることで、伝達関数の係数を定める。つまり、親ノード  $y_4$  は、子ノードから得られた予測値を入力として受け取り、他方への入力  $x_{t-2}$  を用いて上記と同様の計算を行い、その予測値を親ノードに伝える。そして最終的に、 $y_2$  の出力として、より予測対象に近い予測値を得ることができる。

## 2.2 評価関数

GP では、まず適当な初期状態を設定し、上述の木構造による予測モデル  $\{f_i\}_{i=1}^n$  を多数 ( $N$  個) 生成する。そして、次式にて説明する評価関数を用いて各予測モデルのデータへの適

合度を評価し、次世代に残す予測モデルや交差、突然変異などの遺伝的操作の対象となる予測モデルを選択する。GP の要素となる各予測モデルの評価関数として、Rational Average Error (RAE) に基づく Generalized Cross-Validation (GCV)<sup>9)</sup> を用いた。

$$RAE = \sum_{i=1}^T (x_t - f(\mathbf{x}_{t-1}))^2 / \sum_{t=1}^T (x_t - x_{t-1})^2 + \lambda \sum_{j=1}^A a_j^2, \quad (2)$$

$$GCV = RAE / (1 - A/T)^2.$$

ここで、 $A$  は予測モデルを構成する伝達関数の係数の総数とする。学習データを入力した時の各予測モデルの出力に対する GCV を計算し、GCV の値の小さいものをより適合度の高い個体とみなし、適合度の高い個体とそれらの間での交差、突然変異により生成された個体を次世代に残す。こうした進化的世代交代を繰り返すことで、より適合度の高い個体を探索する。なお、他の多くの時系列解析手法と同様に、GP による時系列予測に先立ち、扱う時系列データを適切に変換する必要がある。文献 8) において、スムージング処理、対数変換の差 (前期比増減率近似) と、移動平均除去処理  $x_t \leftarrow x_t - \frac{1}{l} \sum_{i=t-l+1}^t x_i$  による前処理を比較しており、本研究では文献 8) において最も性能が良いと結論づけられている移動平均除去処理を  $l = 5$  とし、行ったデータを扱うことにする。

## 3. 予測モデルの重み最適化

前節で紹介した GP によって構成した複数の株価予測モデル  $f_i(\mathbf{x}_{t-1})$  の加重平均を用いて、時刻  $x_t$  の予測を行うことを考える。この様に多数の予測モデルの出力を平均した予測モデルの構築手法の代表的なものとして Bagging<sup>11)</sup> があるが、Bagging では図 2(a) のように等しい重みで予測値の平均をとる。一方、提案手法では、文献 10) で提案された手法を用いて予測モデルの重みを最適化し、予測モデルの出力の分散を安定化する手法を提案する。この重み最適化により、予測器の分布は図 2(b) のように真の値近くに集中して分布することが期待される。

### 3.1 正規化項付きのクロスエントロピー推定量

確率密度関数  $f(x), g(x)$  に対して、クロスエントロピーは

$$H(f, g) = E_f [-\log g(x)] = - \int f(x) \log g(x) dx \quad (3)$$

で定義される量であり、密度関数  $f(x)$  を持つ分布からデータ  $\mathcal{D}_x = \{x_i\}_{i=1}^n$  が観測されたとすると、 $f(x)$  に関する期待値を経験期待値で置き換えたクロスエントロピーは、モデル  $g(x)$

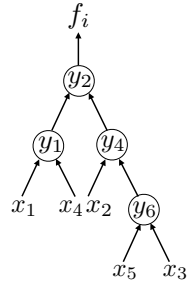


図1 木構造  
Fig.1 Tree structure

表1 伝達関数  
Table 1 Transfer function

Set of transfer polynomials	
$y_1(u, v) = a_0^{(1)} + a_1^{(1)}u + a_2^{(1)}v + a_3^{(1)}uv + a_4^{(1)}u^2 + a_5^{(1)}v^2$	
$y_2(u, v) = a_0^{(2)} + a_1^{(2)}u + a_2^{(2)}uv + a_3^{(2)}u^2 + a_4^{(2)}v^2$	
$y_3(u, v) = a_0^{(3)} + a_1^{(3)}v + a_2^{(3)}uv + a_3^{(3)}u^2 + a_4^{(3)}u^2$	
$y_4(u, v) = a_0^{(4)} + a_1^{(4)}u + a_2^{(4)}v + a_3^{(4)}u^2 + a_4^{(4)}v^2$	
$y_5(u, v) = a_0^{(5)} + a_1^{(5)}uv + a_2^{(5)}u^2 + a_3^{(5)}v^2$	
$y_6(u, v) = a_0^{(6)} + a_1^{(6)}u + a_2^{(6)}uv + a_3^{(6)}u^2$	
$y_7(u, v) = a_0^{(7)} + a_1^{(7)}u + a_2^{(7)}uv + a_3^{(7)}v^2$	
$y_8(u, v) = a_0^{(8)} + a_1^{(8)}v + a_2^{(8)}uv + a_3^{(8)}u^2$	
$y_9(u, v) = a_0^{(9)} + a_1^{(9)}v + a_2^{(9)}uv + a_3^{(9)}v^2$	
$y_{10}(u, v) = a_0^{(10)} + a_1^{(10)}u + a_2^{(10)}v^2 + a_3^{(10)}uv$	
$y_{11}(u, v) = a_0^{(11)} + a_1^{(11)}u + a_2^{(11)}v + a_3^{(11)}u^2$	
$y_{12}(u, v) = a_0^{(12)} + a_1^{(12)}u + a_2^{(12)}v + a_3^{(12)}v^2$	
$y_{13}(u, v) = a_0^{(13)} + a_1^{(13)}u^2 + a_2^{(13)}v^2$	
$y_{14}(u, v) = a_0^{(14)} + a_1^{(14)}u + a_2^{(14)}v^2$	
$y_{15}(u, v) = a_0^{(15)} + a_1^{(15)}v + a_2^{(15)}u^2$	

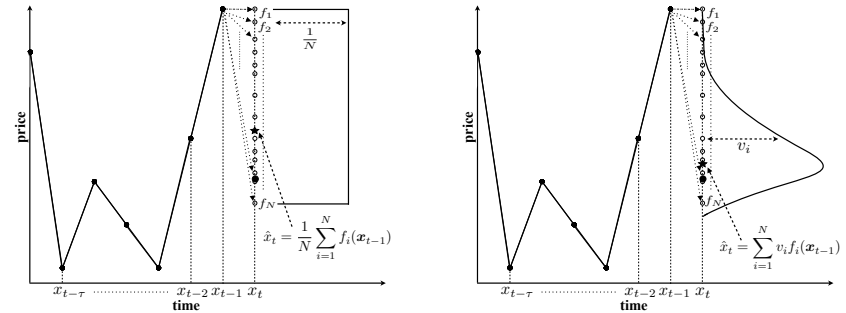
のもとでのデータ  $x_i$  の負の尤度の期待値を表す。つまり、クロスエントロピーが小さいほど、データ集合  $\mathcal{D}_x$  が従う分布の確率密度関数が  $g(x)$  に近いということである。今、それぞれ密度関数  $f(x), g(y)$  に従うと考えられる2組のデータセット  $\mathcal{D}_x = \{x_i\}_{i=1}^n, \mathcal{D}_y = \{y_j\}_{j=1}^m$  が与えられており、特にデータ  $\mathcal{D}_y$  の各要素  $y_j$  に重みが付与されている場合、つまり

$$\mathcal{D}_{yv} = \{\mathcal{D}_y, \mathcal{V}\} = \{(y_j, v_j)\}_{j=1}^m, \sum_{j=1}^m v_j = 1, v_j \geq 0 \quad (4)$$

を考える。こうしたデータに対して、 $y$  の関数  $F(y)$  の経験期待値を  $\sum_{j=1}^m v_j F(y_j)$  で定義する。文献 10) において、このような重み付きデータに対する効率的なクロスエントロピーの推定量

$$\hat{H}(\mathcal{D}_x, \mathcal{D}_{yv}) \propto \sum_{i=1}^n \sum_{j=1}^m v_j \log \|y_j - x_i\| \quad (5)$$

が提案された。このクロスエントロピーを重み  $\mathcal{V}$  に関して最小化することで、重み付きデータ  $\mathcal{D}_{yv}$  の分布を経験分布に近づける。このとき、限られたデータを用いて実質的には尤度の最大化を行うことになり、データへの過適合が生じる可能性がある。そこで、重



(a): 予測値の平均を用いた予測

(b): 予測値の加重平均を用いた予測

図2 提案する株価予測手法。

Fig. 2 Proposed stock price prediction method

み  $\mathcal{V}$  が多項分布のパラメタとみなせることに着目し、重み  $\mathcal{V}$  がデータ  $\mathcal{D}_y$  とは独立にパラメタ  $\beta = \{\beta_j\}_{j=1}^m$  を持つ Dirichlet 分布  $h(v; \beta) \propto \prod_{j=1}^m v_j^{\beta_j - 1}$  に従うとして、クロスエントロピーの推定量に重みの分布のエントロピーを加えた正則化項付きの推定量  $\hat{H}(\mathcal{D}_x, \mathcal{D}_{yv}) - \log h(v; \beta) = \hat{H}(\mathcal{D}_x, \mathcal{D}_{yv}) - \sum_{j=1}^m (\beta_j - 1) \log v_j$  を用いることで、より安定した  $\mathcal{V}$  に関する最適化が可能となる。ここでは特定の重みの大小に関する事前知識は無いものとして、全ての  $j$  に関して  $\beta_j = \mu + 1$  と置く。

### 3.2 予測モデルと予測対象とのクロスエントロピー最小化

前小節のクロスエントロピー推定において時系列データの一点  $x_t$  を  $\mathcal{D}_x^{(t)}$  として、時刻  $t$  より前の  $\tau$  時刻分のデータ  $x_{t-1} = \{x_{t-\tau}, \dots, x_{t-1}\}$  を入力とした GP による  $x_t$  の予測値を  $\mathcal{D}_y^{(t)} = \{\hat{x}_t = f_j(x_{t-1})\}_{j=1}^m$  とすると、 $\{\mathcal{D}_x^{(t)}, \mathcal{D}_y^{(t)}\}_{t=\tau+1}^T$  である  $T - \tau$  個のデータセット対のクロスエントロピーを推定することとなる。従って、 $x_t$  と  $\{\hat{x}_t = f_j(x_{t-1})\}_{j=1}^m$  との正則化項付きのクロスエントロピー推定量は定数倍を除いて  $\sum_{j=1}^m v_j \|f_j(x_{t-1}) - x_t\| - \mu \sum_{j=1}^m (\beta_j - 1) \log v_j$  であり、これを時系列データすべての時刻に対して平均したものを

$$J(\mathcal{V}) = \frac{1}{T - \tau} \sum_{j=1}^m \sum_{t=\tau+1}^T v_j \log \|f_j(x_{t-1}) - x_t\| - \mu \sum_{j=1}^m (\beta_j - 1) \log v_j \quad (6)$$

において、この関数  $J(\mathcal{V})$  を最小化の目的関数とする。最適化の手法は任意であるが、本研究では、勾配法を用いて  $J(\mathcal{V})$  を最小化する。正則化パラメタ  $\mu$  は、予備的な実験により  $\mu = 1.0 \times 10^{-6}$  と定めた。

表 2 GP の各種設定  
Table 2 Setup of GP

個体数	100	最小ノード数	3
世代数	10	最大ノード数	40
交差率	1.0	選択法	トーナメント方式 (size:2)
突然変異率	0.1	式 (2): $\lambda$	0.00001

## 4. シミュレーション

提案する予測モデルの重み最適化による予測値分布の分散低減手法の効果を検証するため、人工データと実データを用いた実験を行う。GP の各種設定を表 2 に示す。

### 4.1 人工データによる検証

まず、人工的に生成した予測対象データと予測モデルに対して、式 (6) の最小化を行うことで、予測モデルに適切な重みを設定出来ることを確認する。次に、重みの最適化により、複数数のモデルの加重平均による予測モデルを構築し、この加重平均モデルが予測精度を落とすことなく分散を低減出来ることを確認する。

#### 4.1.1 重みの最適化実験

予め 3 個の異なる予測モデルを、過去 20 時刻分のデータと表 1 に示した伝達関数をランダムに組み合わせることにより構築する。以下に 3 個の予測モデルを示す。入力値には、実データによる検証 (4.2) で用いる日経平均株価 7000 点を用いる。

$$\begin{aligned}
 f_{1,t}(x_{t-4}, x_{t-17}) &= a_1^{(1)} + a_2^{(1)} y_3(x_{t-4}, x_{t-17}) + a_3^{(1)} y_3(x_{t-4}, x_{t-17}) y_5(x_{t-4}, x_{t-17}) \\
 &\quad + a_4^{(1)} y_3(x_{t-4}, x_{t-17})^2 + a_5^{(1)} y_5(x_{t-4}, x_{t-17})^2, \\
 f_{2,t}(x_{t-5}, x_{t-11}) &= a_1^{(2)} + a_2^{(2)} y_6(x_{t-5}, x_{t-11}) + a_3^{(2)} y_6(x_{t-5}, x_{t-11}) y_{12}(x_{t-5}, x_{t-11}) \\
 &\quad + a_4^{(2)} y_6(x_{t-5}, x_{t-11})^2 + a_5^{(2)} y_{12}(x_{t-5}, x_{t-11})^2, \\
 f_{3,t}(x_{t-6}, x_{t-13}) &= a_1^{(3)} + a_2^{(3)} y_{11}(x_{t-6}, x_{t-13}) + a_3^{(3)} y_{11}(x_{t-6}, x_{t-13}) y_{15}(x_{t-6}, x_{t-13}) \\
 &\quad + a_4^{(3)} y_{11}(x_{t-6}, x_{t-13})^2 + a_5^{(3)} y_{15}(x_{t-6}, x_{t-13})^2.
 \end{aligned}$$

各係数  $a_j^{(i)}$  は、各予測モデルの出力値が、それぞれ異なる擬似的な株価を表現できるように適切な設定をしている。つまり、全予測モデルから異なる 3 種類の出力値を得る。そして各時刻において、その 3 種類の出力値の中から 1 つの出力値を選ぶ。この時、出力値の選択は、予め各予測モデルに対して設定した重みの割合に従う。各時刻において選ばれた一連の出力値を予測対象データとする。  $f_1$  に対する重みを  $v_1 = 0.1$ ,  $f_2$  に対する重みを

$v_2 = 0.3$ ,  $f_3$  に対する重みを  $v_3 = 0.6$  と設定した場合、生成される予測対象データは、例えば  $\mathcal{D}_f = \{f_{3,t}, f_{3,t+1}, f_{3,t+1}, f_{3,t+2}, f_{2,t+3}, f_{2,t+4}, f_{3,t+5}, f_{3,t+6}, f_{2,t+7}, f_{1,t+8}, \dots\}$  となり、0.1, 0.3, 0.6 の割合で 1 つの予測モデルの出力値が選ばれたことになる。この様にして生成された予測対象データと予測モデルに対して、式 (6) の最小化を行い適切な重みを得る。基本となる 3 個の予測モデルに対して予め設定した 3 パターンの重みと、最適化の結果得られた重みを表 3 に示す。表 3 から、基本となるモデルそれぞれに予め設定した重みに非常に近い値が最適化結果として得られていることが分かる。

#### 4.1.2 分散の安定化

重みの最適化実験と同様の設定で、予め 100 個の予測モデルを構築する。入力値も同様に実データによる検証 (4.2) で用いる日経平均株価 7000 点を用いる。ここでは、各時刻において、100 個の予測モデルから得られる出力値の期待値に対して、出力値の標準偏差を  $\sigma$  とし、 $[-\sigma/3, \sigma/3]$  の範囲で一様乱数を加えることで得られる一連のデータを予測対象データとする。つまり、予測対象データが予め全出力値の期待値  $\pm 1 \times$  標準偏差までの範囲にすべて含まれている状態となる。

この様にして生成された予測対象データと予測モデルに対して、式 (6) の最小化を行った結果、全時刻における分散を平均すると重み一様の場合には 11.45、重み最適化後には 7.05 であり、両者を比較すると約 38.4% の分散の低減が達成された。また、分散低減後も期待値  $\pm 1 \times$  標準偏差までの範囲に予測対象データはすべて含まれおり、予測精度は低下していないといえる。図 3 に、重みを最適化したことによる期待値  $\pm 1 \times$  標準偏差までの範囲の変化を示す。

### 4.2 実データによる検証

実データには、日経平均株価 (2009/08/11~2009/12/31) の 1 分足の終値を扱い、20 時刻前までのデータを用いて 1 分後の株価の予測を行う。全データ数は 18000 点であり、予測モデルの構築 (GP) に 3000 点、重みの最適化に 5000 点、テスト期間として 10000 点を用いた。まず、各予測モデルの平均予測誤差 (Mean Prediction Error:MPE) に対して、最適化された重みの散布図を図 4 に示す。図 4 より、重み最適化の結果として誤差の大きい予測モデルに対して小さい重みが、誤差の小さな予測モデルには大きい重みが与えられることが分かる。次に、重み最適化の予測精度への影響と、予測分布の分散への影響を考察する。

#### 4.2.1 予測精度についての実験

以下の 3 通りの設定で株価の予測実験を行う。

- (1) 全予測モデルの中で最も精度の高いモデルのみを用いた場合 (Best Model)

表 3 設定した重みと最適化結果  
Table 3 Defined weights for element models and optimized weights

$v_1$	$v_2$	$v_3$	$f_1$	$f_2$	$f_3$
0.2	0.3	0.5	0.21	0.29	0.50
0.5	0.2	0.3	0.48	0.20	0.32
0.3	0.5	0.2	0.29	0.50	0.21

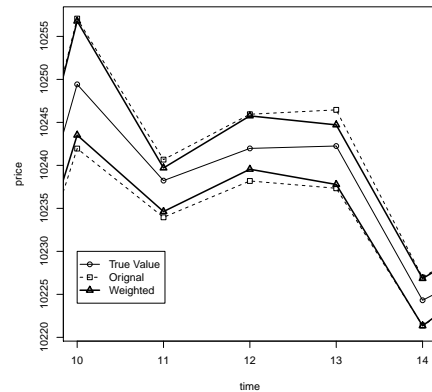


図 3 分散の低減  
Fig. 3 Cut down dispersion

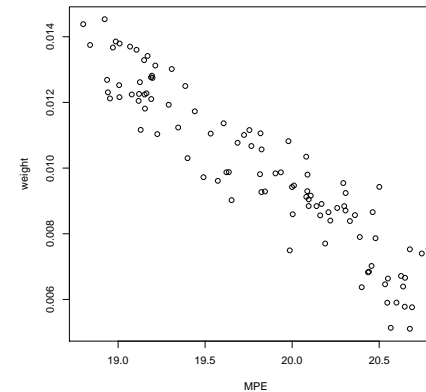


図 4 誤差に対する重みの分布  
Fig. 4 Correspondence between MPE and weight

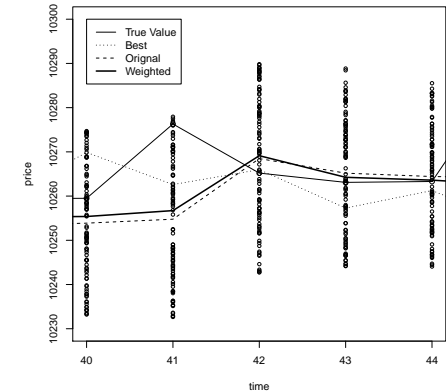


図 5 予測結果  
Fig. 5 Result of prediction

(2) 全予測モデルの単純平均を用いた場合 (Original Model)  
 (3) 全予測モデルの重みを最適化した加重平均を用いた場合 (Weighted Model)  
 以上 3 パターンにおける比較実験を行い、平均予測誤差 (MPE), 及び次期の値が現在の値よりも上昇または下降を予測した的中率 [%](Hit Rate:HR) に関する比較結果を表 4 に示す。表 4 の結果では、交差検証を 5 回行った上での平均値と標準偏差 (sd) を示している。また、全モデルによる予測値を含む上記の 3 パターンによる時系列データを図 5 に示す。表 4 及び図 5 より、単一の最も精度の高いモデルを用いるよりも多数のモデルの単純平均を用いる方が平均予測誤差、的中率いずれも向上することが分かる。さらに、重みを最適化した予測モデルの出力値の加重平均を予測に用いることで、予測性能はさらに向上することが確認できる。重みを一様にした場合と最適化した場合の平均予測誤差、及び的中率の差異は、t 検定により 5% 水準で有意である。

#### 4.2.2 予測分布についての実験

予測分布を形成した場合の全予測値の期待値  $\pm 1 \times$  標準偏差までの範囲に実測値が含まれている割合、及び分布の分散の比較を行う。

- (1) 全予測モデルの重みを一様とした場合 (Original Model)
- (2) 全予測モデルの重みを最適化した場合 (Weighted Model)

以上 2 パターンにおける比較実験を行い、予測値の分散 (Variance), 及び実測値が期待値  $\pm 1 \times$  標準偏差までの範囲に含まれる割合 [%](Including) の比較結果を表 5 に示す。また、表 5 の結果では、交差検証を 5 回行った上での平均値と標準偏差 (sd) を示している。そして、重みを設定したことにより得られた予測分布の時系列データを図 6 に示す。表 5 より、予測モデルに対する重みを最適化することで、予測分布の期待値  $\pm 1 \times$  標準偏差までの範囲に実測値が含まれる割合はほぼ一定に保った上で、予測分布の分散を低減出来ていることが分かる。これより、より安定した予測が可能になると考えられる。なお、これも 5% 水準の t 検定で有意差がある結果である。

## 5. おわりに

本研究では非線形回帰問題において、複数の予測モデルに適切な重みを与えることで、予測分布の分散を予測精度を落とすことなく低減する手法を提案した。一般的に、予測精度が変わらない場合、より分散の小さい分布が信頼性の高い予測分布であると考えられる。重みの最適化手法としては、最近提案された情報量の推定量に基づくクロスエントロピーの最小化を採用した。

本研究では、基本となる予測モデルとして GP を採用したが、基本となる予測モデルは様々

表 4 予測精度の比較  
Table 4 Predictive accuracy

	MPE (sd)	HR (sd)
Best Model	18.88 (0.078)	56.82 (0.072)
Original Model	15.45 (0.089)	58.11 (0.150)
Weighted Model	15.32 (0.088)	58.31 (0.183)

表 5 予測分布の比較  
Table 5 Predictive distribution

	Variance (sd)	Including (sd)
Original Model	174.89 (1.230)	63.17 (0.073)
Weighted Model	170.44 (1.295)	63.38 (0.171)

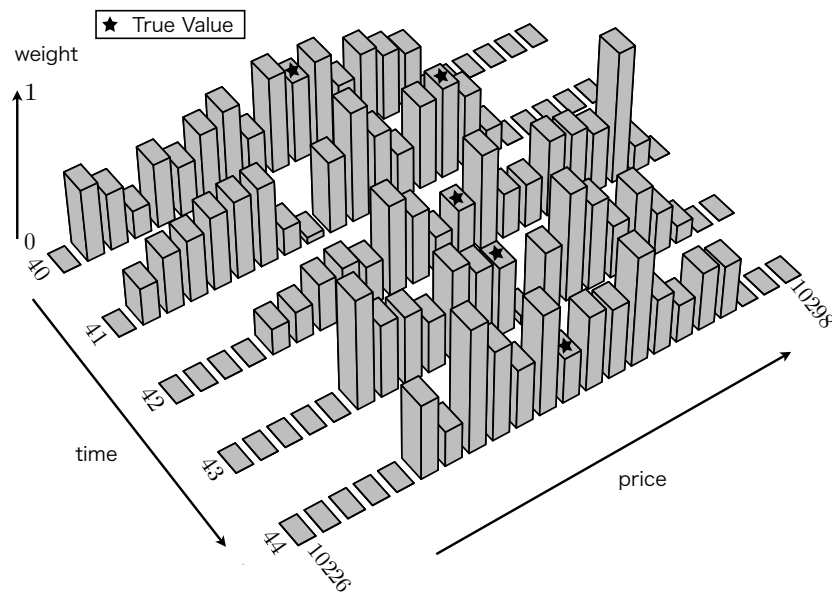


図 6 予測分布  
Fig.6 Result of prediction

考えられる。特に、表 2 に示したように、本研究で利用した GP の学習回数は 10 回としており、一般的な GP の学習回数と比較すると非常に少ない。これは、予測モデルの多様性を維持することが目的である。アンサンブル学習に関する研究でよく知られているように、多数のモデルを平均して予測を行う場合には、基本となる予測モデル群は十分な多様性を有する必要がある。今後は入力変数として様々な経済変数を取り入れるとともに、予測モデルの個数を増やすことでさらに予測モデルに多様性を持たせ、より良い予測分布を構築していくことに取り組む。なお、実際に構築した予測分布を用いて、よりリスクの少ない株取引を行うことが最終的な研究目的である。図 6 に示したように、本研究における提案手法によって得られる各時刻の株価の予測値の分布は、一般には単峰性とは限らず、よく仮定されるような正規分布とは異なる、こうした予測分布を活用した投資手法を提案することも今後の課題の一つである。

謝辞 本研究の一部は科研費 22800067 の助成を受けたものである。

### 参 考 文 献

- 1) 刈谷武昭 ほか: 経済時系列の統計, 岩波書店 (2003).
- 2) 森平爽一郎 ほか: ファイナンスへの確率解析, 朝倉書店 (2000).
- 3) Y.Yoon and G.Swales: Predicting stock price performance: A neural network approach, In *Proc. Twenty-Fourth Annual Hawaii International Conference on System Sciences*, (1991).
- 4) W.Huang and Y.Nakamori and S.Wang: Forecasting stock market movement direction with support vector machine, *Computers and Operations Research*, Vol.32, No.10, pp.2513-2522 (2005)
- 5) S.Mahfoud and G.Mani: Financial forecasting using genetic algorithms, *Applied Artificial Intelligence*, Vol.10, pp.543-565, (1996)
- 6) H.Iba and T.Sasaki: Using Genetic Programming to Predict Financial Data, In *Proc. 1999 Congress on Evolutionary Computation*, (1999)
- 7) 伊庭齊志: 遺伝的プログラミング入門, 東京大学出版会 (2001).
- 8) H.Iba and N.Nikolaev: Genetic programming polynomial models of financial data series, In *Proc. the 2000 Congress on Evolutionary Computation* (2000)
- 9) T.Hastie and R.Tibshirani and J.Friedman: *The Elements of Statistical Learning - Data Mining, Inference, and Prediction - (2nd edition)*, Springer, (2009)
- 10) 日野英逸 ほか: 分位点に基づく重み付きデータの情報量推定手法とその応用, 第三回 IBISML 研究会 (2010).
- 11) L. Brieman: Bagging Predictors, *Machine Learning*, Vol.24, No. 2, pp.123-140 (1996).