アンサンブル予測偏差モーメントを用いる時系列予測の性能評価

九州工業大学　○小野航平　黒木秀一　西田健

Experimental Analysis of Moments of Predictive Deviations as Ensemble Diversity Measures for Model Selection in Time Series Prediction

Kohei Ono, Shuichi Kurogi, and Takeshi Nishida, Kyushu Institute of Technology

Abstract: This paper presents an experimental analysis of moments of predictive deviations as measures of ensemble diversity to estimate the performance of time series prediction for model selection. As an extension of the ambiguity decomposition of bagging ensemble, we decompose the fourth power of ensemble prediction error and examine the effect of the moments of predictive deviations of ensemble members to the ensemble prediction error. We analyze the result of numerical experiments and show some properties and the effectiveness of the moments of predictive deviations.

1. まえがき

本稿では，時系列予測の性能を推定してより良い予測モデルを選択するため，アンサンブル予測の多様性測度として予測偏差モーメントを用いる手法について実験解析を行う．ここでアンサンブル予測に関する不一致の度合いを表す多様性測度は，アンサンブル学習に応用され，その有効性が示されている．例えば既知の目標値を持つ訓練データセットを用いた負相関学習は，予測誤差の最小化と予測誤差の共分散の最大化のトレードオフの適切な調整により，より良い性能が得られることが示されている．しかし，この学習法は目標値が未知の場合の予測性能評価に用いることはできない．そこでアンサンブル予測の誤差の 4 乗を分解し，アンサンブル予測における予測偏差モーメントが予測誤差に及ぼす影響を調査した．次の章ではバギングの記法について示し，バギングアンサンブル予測の誤差の分解と予測偏差モーメントの導入に続いて，マルチステップの時系列予測について示す．3 章では数値実験の結果と本解析の有効性について示す．

2. バギング，多様性測度，時系列予測

2.1 回帰問題の為のバギング

$D^n \overset{\Delta}{=} \{(x_i, y_i) | i \in I^n\}$ を訓練データセットとする．x_i, y_i はそれぞれ入力ベクトルと目標値を示し，また，$I^n=\{1,\ldots,n\}$ である．ここで式(1)のような関係が与えられるとする．

$$y_i \overset{\Delta}{=} r_i + e_i = r(\boldsymbol{x}_i) + e_i, \qquad (1)$$

$r_i=r(\boldsymbol{x}_i)$ は x_i の非線形な目標関数，e_i は平均が 0 の誤差を表す．

次にバギングについて定式化する．$D^{n\delta^{\#},j}$ を $n\alpha$ 個の要素を含む j 番目のバグとする．この中の要素は訓練データセットからランダムな復元抽出により生成される．ここで α はバグサイズ比を示し，さらに $j=\{1, \cdots, b\}$ である．ここで多くのアプリケーションでは $\alpha=1$ が使われるが，本研究では汎化性能を向上させるために変数 α を用いることに注意する．複数の学習機械 θ^j を用いて $D^{n\delta^{\#},j}$ を学習し，式(2)によって目標値を推定するバギングが行われる．

$$\hat{y}_i^{\mathrm{bag}} \overset{\Delta}{=} \hat{y}^{\mathrm{bag}}(\boldsymbol{x}_i) \overset{\Delta}{=} \frac{1}{b} \sum_{j \in J^{\mathrm{bag}}} \hat{y}_i^j \equiv \left\langle \hat{y}_{ij} \right\rangle_{j \in J^{\mathrm{bag}}} \qquad (2)$$

$\hat{y}_i^{bag} \overset{\Delta}{=} \hat{y}^{bag}(\boldsymbol{x}_i)$ は j 個の学習機械 θ^j による予測を示す．またかぎ括弧は平均を示し，右下の添え字は平均をとる範囲を表す．

2.2 誤差分解と予測偏差モーメント

バギングアンサンブルの誤差を解析するためにバイアス分散の分解と分解の曖昧さについて調査した．まず，各予測を $\hat{y}_i^j = r_i + \beta_i + \varepsilon_i^j$ とする．β はバイアスを表し，ε は予測偏差を表す．そして全てのバグの予測と訓練データとの平均二乗誤差は，

$$\begin{aligned} \left\langle (\hat{y}_i^j - y_i)^2 \right\rangle_j &= \left\langle (\beta_i + \varepsilon_i^j - e_i)^2 \right\rangle_j \\ &= (\beta_i)^2 + \left\langle (\varepsilon_i^j)^2 \right\rangle - 2\beta_i e_i + e_i^2, \end{aligned} \qquad (3)$$

と分解される．さらに，一般に汎化誤差と呼ばれるバギング予測と真値の二乗誤差は，

$$\left(\hat{y}_i^{bag} - r_i\right)^2 = \left(\beta_i\right)^2 = \left\langle \left(\hat{y}_i^j - y_i\right)^2 \right\rangle_j - \left\langle \left(\varepsilon_i^j\right)^2 \right\rangle + 2\beta_i e_i - e \quad (4)$$

と分解される．ここで式(3)はバイアスーバリアンス分解(bias-variance decomposition)に対応し，式(4)は曖昧さ分解(ambiguity decomposition)に対応する．ここで ε_i^j の分散の項は多様性測度のひとつであり，曖昧さ(ambiguity)と呼ばれる．この曖昧さの分解の式から，他の項が一定ならば，大きな分散は汎化誤差を小さくすると仮定される．さらに，未知の値を予測するときに分散の項のみを得ることができるので，分散は汎化誤差を推定するのに有効であることが期待できる．しかし後述の実験結果が示すように，分散と汎化誤差は無関係である．これは，第 1 項が分散の影響を打ち消すからであると考えられる．そこで次のように誤差の 4 乗を分解してみる．

$$\begin{aligned}\left(\hat{y}_i^{bag} - r_i\right)^4 &= C - \left\langle (\varepsilon_i^j)^4 \right\rangle_j - 4\left(\hat{y}_i^{bag} - y_i\right)\left\langle (\varepsilon_i^j)^3 \right\rangle_j \\ &\quad - 6\left(\hat{y}_i^{bag} - y_i\right)^2 \left\langle (\varepsilon_i^j)^2 \right\rangle_j \end{aligned} \quad (5)$$

C は ε を陽に含まない項の和を示す．ここで $\hat{y}_i^{bag} - y_i = \beta - e_i$ の項は未知の y_i を含まないことに注意する．そして右辺の C が一定のとき，左辺を減少させるためには $\left(\varepsilon_i^j\right)^4, \left(\varepsilon_i^j\right)^2$ の両方が大きくなり，$\left|\left(\varepsilon_i^j\right)^3\right|$ は対応する項が負（正）のとき大きく(小さく)なればよい．

さて，これらの項間の依存関係を除いて汎化誤差を評価するため，予測偏差モーメント，すなわち歪度 S_i，尖度 K_i，および分散 V_i を考える．

$$V_i \equiv \sigma_i^2 \equiv \left\langle \left(\varepsilon_i^j\right)^2 \right\rangle_j, S_i \equiv \frac{\left\langle \left(\varepsilon_i^j\right)^3 \right\rangle_j}{\sigma_i^3}, K_i \equiv \frac{\left\langle \left(\varepsilon_i^j\right)^4 \right\rangle_j}{\sigma_i^4}. \quad (6)$$

全てのテストデータを予測するために平均分散 MV $\left\langle V_i \right\rangle_j$，歪度の絶対値平均 MAS $\left\langle |S_i| \right\rangle_j$，平均尖度 MK $\left\langle K_i \right\rangle_j$ を用いる．S ではなく $|S|$ を用いるのは 3 乗の項の未知の極性を合わせるためである．

2. 3　時系列予測

上記の解析は時系列多段予測の性能を推定するために，以下のように利用する．

離散時間 t＝0,1,2,…に対する実数値 $y(t)$の時系列を $y_{t:n}$=$y(t)y(t+1)\ldots y(t+n+1)$とする．与えられた時系列を $y_{\{t_g,n_g\}}$，に対し，それより後の時系列 $y_{\{t_p,n_p\}}$を予測することを考える．ここで $t_p \geq t_g + t_n$ である．この問題を解決するために $y_t = r(x_t) + e_t$ (式 1)に $y_t = y(t), x_t = (y(t-1), y(t-2), \ldots, y(t-k))^T$ を代入したものを用いる．ここで k は埋め込み次元を示し，適切なものを選ばなければならない．また学習と予測は上述の回帰問題として定式化できる．さらに多段予測は $\hat{y}_t = \hat{y}^{bag}(\hat{\boldsymbol{x}}_t)$により逐次的に実行できる．ここで $\hat{\boldsymbol{x}}_t = (x_{t1}, x_{t2}, \cdots, x_{tk})$ であり $x_{tj} = y_{t-j}(t - j < t_g)$ または $x_{tj} = \hat{y}_{t-j}(t - j \geq t_g)$ において t＝t_p,t_p＋1,…である．

3. 数値実験

3.1　実験の設定

カオス時系列として図 1 に示す Lorenz の時系列を用いる．

バギングの学習機械の訓練として $y_{0:2000}$ を用いる．また，初期入力ベクトル x_{tp}=($y(t_p$-1),…,$y(t_p$-$k)$)の $y_{tp:np}$ の多段予測は，t_p=2000+100i(i=0,1,2,…,29) を予測開始時刻とする．また予測範囲は n_p=1,10,50,100 とする．最後に，バギングの予測偏差モーメントの解析を行う．ここで，本実験における訓練データと予測データの関係は，一般的に t_p=t_g+n_g を用いるものとは異なることに注意する．しかし，下に示すようにモーメントのいくつかの大きな性質を得ることができた．

学習機械として，CAN2 を用いる．ここでモデルの複雑さはユニット数 N であり，予測関数を近似する区分線形領域の数を表す．問題を解決するために N={20,40,…,300}の中から適切な N を選択し，よりよい予測を実現させたい．埋め込み次元は k=8，バグ数は b=100 を用いる．これは何度かの試行において良い予測性能を示したからである．

3.2　結果と解析

t_p=2000+100i と n_p=1,5,10,100 の組み合わせは全部で 120 通りある．全てにバギング予測を実行し全ての N に対して MSE と予測偏差モーメント即ち MV，MS，MK を得た．実験結果を図 2 に示す．　ここで N_{MSE}，N_{MS}，N_{MK} は各 t_p と n_p のすべての N の中でそれぞれ最小の MSE，最大の MS，最大の MK を実現した N を表す．

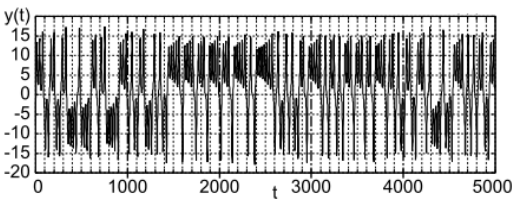

Fig.1. Lorenz time series $y(t)$ for t=0,1,2,⋯,4999.

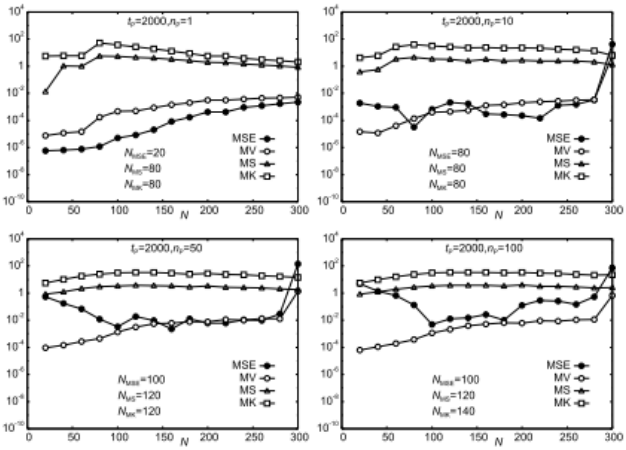

Fig.2. Experimental results of MSE,MV, MS, and MK vs. N for t_p=2000.

全ての t_p と n_p に対する結果を分析することによりいくつかの性質を得た．まず，N の 20 から 300 までの増加に従い，MS と MK はそれぞれまず増加して最大値に達し，そのあと減少するのに対し，MV は最小値から最大値に増加する．これらの増加と減少は僅かな変動は伴うが単調に変化する．この性質は次のようにして得られると考えられる．基本的に，N はモデルの複雑さを与える区分線形領域の数であるので，その大きさが大きくなると，予測偏差モーメントも大きくなると考えられる．ただし，N が大きくなりすぎると区分線形領域内の訓練データ数が増加しなくなるので全てのモーメントは飽和する．全ての予測偏差モーメントはそれも含めて仮定される．歪度と尖度は分散の値によって正規化されるため，分散もしくは 2 次モーメントが 3 次や 4 次のモーメントよりもさらに減少するとき，減少する可能性がある．MSE または予測誤差の 2 次モーメントの最小化は一般的な学習機械も同じだが CAN2 の学習の目的なので，この性質は尤もらしいかもしれない．

次に図 3 の中で N_{MSE} が N_{MS} や N_{MK} よりも小さくなるときの性質を調査した．図 3 において，N_{MSE} または最適なモデルは開始時刻 t_p や予測範囲 n_p が変化すると大きく変化することがわかる．図 3 より，N_{MSE} のおおよその推定値として，N_{MS} や N_{MK} の半値を用いることで予測性能を評価することが考えられる．結果を図 4 に示す. ここで MSE_{min}, MSE_{MS}, MSE_{MK} はそれぞれ N_{MSE}, N=0.5N_{MS}，N=0.5N_{MK} を用いたときの予測値の MSE である．比較のため，$y_{2000+100inp}$ のときに得られた N=N_{MSE} を $y_{2000+100(i+1):np}$ の予測に適用することによって得られる MSEref を求めた．この方法は，t_p の増加に対する最適モデルの変化が連続であるとする仮定に基づいて得られる．この仮定は時系列のモデル選択によく用いられる holdout 法でも用いられていると考えられる．図 4 において n_p=100 のとき，MSE の平均の計算から 10 よりも大きな MSE を生成した N を用いた予測は除外することに注意する．なぜなら 10 を超えた MSE は平均の値を支配するからである. より詳しく説明するため，図 5 に n_p=100，t_p=2200,2300,2400 の予測を正確に示す．ここで $\hat{y}_{MSEmin}$, $\hat{y}_{MS}$, $\hat{y}_{MK}$, $\hat{y}_{ref}$ はそれぞれ N_{MSE}, N_{MS}, N_{MK}, N_{ref} を用いた予測である. tp=2300 の結果より，$\hat{y}_{MS}$, $\hat{y}_{MK}$, $\hat{y}_{ref}$ による予測誤差は t=2350 を超えてから急激に増加する．この誤差の値は非常に大きいため，平均値を支配してしまう．図 4 より，MSE_{MS} と MSE_{MK} の平均値は全ての n_p の範囲において MSE_{ref} よりも小さくなることがわかる．この結果は予測偏差モーメントの有用性を示している．またこの手法は特定の予測開始時刻 t_p や予測範囲 n_p に対して機能することを保証することはできないが，全体の平均としてはうまく機能することが期待できるということを示している．

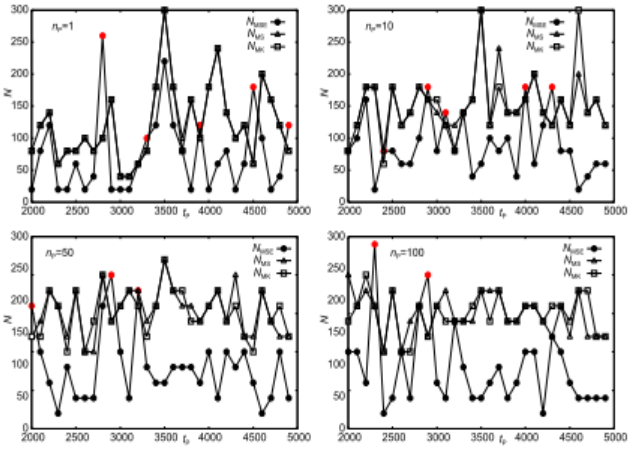

Fig.3. Experimental results of N_{MSE} , N_{MS} and N_{MK} for t_p=2000+100i (i=0,1⋯,29).

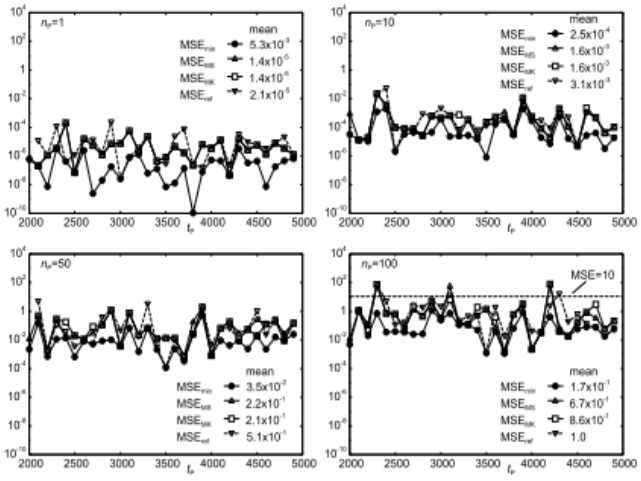

Fig.4. Experimental results of MSE obtained by model selection methods.

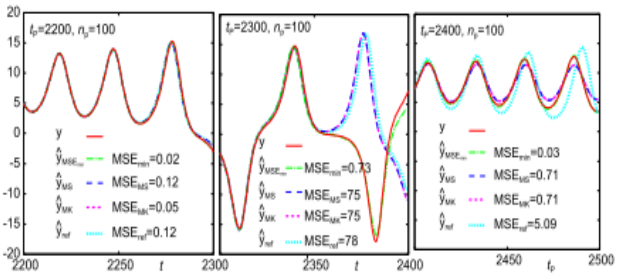

Fig.5. Experimental results of predictions obtained by model selection methods.

4. 結論

本稿では，時系列予測のモデル選択に対して，アンサンブル予測の多様性測度としての予測偏差モーメントが利用できることを示すために解析を行った．バギングの予測誤差の 4 乗から，アンサンブル予測誤差に対する予測偏差モーメントの影響を示した．数値実験により，モーメントのいくつかの性質を示し，モデル選択のための知見を利用した．本手法の有効性は平均的であり，任意の開始時刻と予測範囲に対して保証することはできない．しかし，この問題はカオス時系列の長期予測不能性，すなわち指数的に増加する予測誤差に関係がある可能性があり，解決することは困難である．今後の課題として，この問題に対して他の時系列を用いて調査を行う．

参考文献

1. Brown, G., Wyatt, J., and Tino, P.: Managing diversity in regression ensembles, J. Mach. Learn. Res., Vol. 6, pp.1621–1650, (2005)
2. Chen, H.: Diversity and Regularization in Neural Network Ensembles, PHD thesis, University of Birmingham (2008)
3. Ono, K., Kurogi, S., Nishida, T.: Moments of predictive deviations as ensemble diversity measures to estimate the performance of time series prediction, Proc. ICONIP 2012, Part V, LNCS 7667, pp.59–66. Springer, Heidelberg (2012)
4. Breiman, L.: Bagging predictors, Machine Learning, Vol. 26, no. 2, pp.123–140 (1996)
5. Kurogi, S.: Improving generalization performance via out-of-bag estimate using variable size of bags, J. Japanese Neural Network Society, Vol. 16, no. 2, pp.81–92 (2009)
6. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model se-lection, Proc. of the Fourteenth International Conference 18 on Artificial Intelligence (IJCAI), pp.1137–1143(1995)
7. Efron, B. and Tbshirani, R.: Improvements on cross-validation: the .632+ bootstrap method, J. American Statistical Association, Vol. 92, pp.548–560 (1997)
8. Aihara, K.: Theories and applications of chaotic time series analysis, Sangyo Tosho, Tokyo (2000)