

日経平均株価予測におけるデータ分割手法の検討

畑仲 優輝 平原 誠
法政大学大学院理工学研究科

1. はじめに

近年、機械学習活用した株価予測モデルが多く存在する[1-2]。NN(ニューラルネットワーク)で学習を行う際には学習データを、パラメータ(重み)を決定するための訓練データと、過学習を防いで汎化能力を高めるための検証データとに分割する。この分割方法によって、モデルの性能が左右され、テストデータに対する予測精度の向上が期待できる。本稿では NN での日経平均株価予測における株価データの分割方法を検討する。

2. 予測モデル

本稿では、3 層の NN を用いて 3 営業日分の日経平均株価データから翌日の日経平均株価の「終値前日比」を予測する。入力に、日経平均株価データの連続 3 営業日分の「四本値前日比」を用いる。ここで四本値前日比とは「始値」、「高値」、「安値」、「終値」それぞれの前日比である。加えて現在の株価が移動平均と比較してどの程度乖離しているかを表す「移動平均乖離率」を用いる。移動平均乖離率は「5 日」、「25 日」、「75 日」、「100 日」、「200 日」間のそれぞれ移動平均を「終値」で割ることで求められる。

3. 学習データの分割手法

学習データの中からランダムに検証データを選択する「ランダム」に加えて、テストデータの直近を選択する「直近」、テストデータの直近 1 件おきに選択する「直近交互」の 3 通りの分割方法を検討する。

4. 実験

四本値前日比のみを用いる計 12(=4×3)入力の四本値モデルと、四本値前日比に加え移動平均乖離率を用いる計 27(=9×3)入力の移動平均追加モデルを構築する。各モデルはコスト関数を平均二乗誤差とし、中間素子数を 32,64,128 の 3 段階で最適化する。

日経平均株価データ[3]を取得し、学習データを 1000 件、テストデータを学習データの 1 営業日後の 1 件として実験を行う。1 営業日分ずつデータセットの日付をずらし、計 100 件(2017 年 11 月 22 日～2018 年 4 月 20 日)の予測を行う。1000 件の学習データの中から N={50,100,200}件を選択して検証データとし、残りを訓練データとする。選択の手法を「ランダム」、「直近」、「直近交互」として、計 9 パターン(=3×3)で実験を行う。また前日の終値をそのまま予測日の終値と予測する「後追いモデル」を比較対象として導入する。

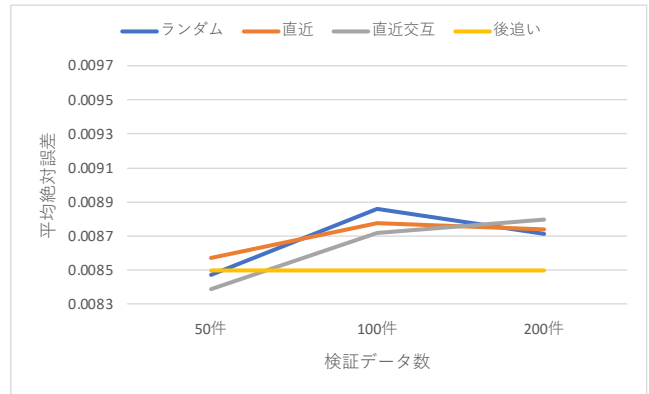


図1. 四本値モデルの平均絶対誤差

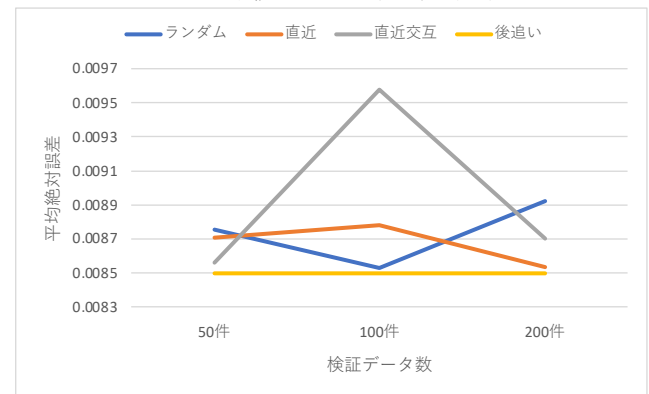


図2. 移動平均追加モデルの平均絶対誤差

5. 実験結果

四本値モデルと移動平均追加モデルについて、9 パターンの検証データの取り方での、テストデータ 100 件に対する平均絶対誤差をそれぞれ図 1 と図 2 に示す。

四本値モデルにおいて検証データ数を 100 件、50 件と少なくしたとき予測精度が向上しており、中でも「直近交互」の精度が他の「ランダム」、「直近」と比較して良く、平均絶対誤差の減少傾向が強かった。

6. 今後の課題

さらなる予測精度向上のため、主成分分析による次元圧縮と中間素子数の変更を検討する必要がある。また今後、試行数を増やし予測精度を観察する必要があると考える。

参考文献

- [1] Mehar Vijh, et al., Procedia Computer Science, vol.167, pp.599-606, 2020.
- [2] Adebiyi AA, et al., J Emerg Trends Comput Inf Sci, vol.3, no.1, pp.1-9, 2012.
- [3] 「investing.com」<https://www.investing.com/>