FDC レポート

team:こんぱいるえらー

2019 年 2 月 22 日

0.1 初めに

私達が今回のコンテストにおいて用いた銘柄は全て日経平均採用銘柄である．日経平均採用銘柄から銘柄を選択した理由は以下の通りである．

- 株価が予測しやすい
- 統計的な予測方法を用いるため

0.1.1 株価が予測しやすい

私達のチームは日経平均採用銘柄は他の銘柄 (マザーズなどで取り扱われている銘柄) と比べて，急激な株価の変動が少ないと考えたので，日経平均採用銘柄の中からポートフォリオに盛り込む銘柄を選出した．

0.1.2 統計的な予測方法を用いるため

私達のチームは，機械学習を用いた開発の経験が一切無かった．なので，他参加チームが実装するであろう機械学習に関する部分で差をつけられてしまうと考えた．よって私達のチームでは，機械学習でカバーしきれないところを統計的な予測方法を用いることによってカバーした．

1 データの種類について

私達が今回使用したデータは主に 2 種類である．

- 過去数年分の株価データ
- 過去数年分の業績データ

1.0.1 過去数年分の株価データ

株価データは 2014〜2019 年までのデータを用いた．https://kabuoji3.com/からスクレイピングを用いて株価データを抽出した．なお抽出した株価データは，/trade/adopted_nikkeiheikin/201n/all_stock_data_with_date に保存している．

1.0.2 過去数年分の業績データ

業績データも株価データを同じく，2014〜2019 年までのデータを用いた．https://shikiho.jp/ からスクレイピングを用いて業績データを抽出した．なお抽出した業績データは，/trade/toushou.csv に保存している．

2 データの前処理について

今回のコンテストでは複数の方法を用いて機械学習を行った．trade.py では、株価の上昇率を大体-1.0〜1.0 に収まるように調整を行った trade2.py では、業績の変動率を大体-1.0〜1.0 に収まるように調整を行った また、sklearn の機能を用いて正規化、標準化、も行った。

3 モデルの説明・工夫について

3.0.1 モデルの説明

今回のコンテストで作成したモデルでは、株価の終値を用いて株価の上昇率を算出し、翌日の株価が前日の終値と比べて上昇しているか否かを one hot 表現を用いて正解ラベルを作成し，機械学習を行った。また，今回のコンテストでは複数のモデルを作成し用いた．1 つめのモデルは，株価の終値を 10 日分学習させ，次の日の株価が上昇するか否かを予想する方法である．2 つめのモデルは，株価の終値だけではなく始値，高値，安値，出来高なども加味し、次の日の株価が上昇するか否かを予想する方法である．

3.0.2 モデルの工夫

モデルの工夫を行った点は主に 2 つある．

- 複数のモデルを用いる
- 統計的な予測方法を織り交ぜる

3.0.3 複数のモデルを用いる

先ほど説明した通り，今回のコンテストでは複数のモデルを用いた．複数のモデルを用いた理由は先ほど説明した通り，他参加チームと比べて私達のチームのモデルを用いた実装経験が大きく劣ると考えたためである．

3.0.4 統計的な予測方法を織り交ぜる

モデルの工夫を行った点の 1 つである複数のモデルを用いるということには一つの大きな欠点が存在する．それは予想出来る範囲がコンテスト開始日のみである．この欠点を埋めるため，私達は統計的な予測方法を用いた．短期的な予想を可能にする機械学習と長期的な予想を可能にする業績を用いた統計的な予測方法を組み合わせることによって，中長期的に株価が上昇する銘柄を選択することができると考えたためである．

4　予測結果に関する考察

複数の学習方法や統計的な予測方法を用いて選出した銘柄は 8 つあった．8 つのうち製薬会社が 2 つ，半導体などの工業関係の企業が 2 つ，そのどちらも行っている企業が一つ．人材関連の企業が 1 つ，保険関連の企業が 1 つ，鉄道関連の企業が一つと様々な分野の企業の銘柄を選出した．日本の輸出の内訳を用いて考察すると，以下のことが考えられた．

- 世界に輸出している品の内，ここ数年の上位 10 位以内に入っている品を製造している企業が選出した企業のう内の過半数を占めている
- 輸出額は年々上昇している傾向にある．

この 2 つの情報から私達のチームでは，ここ数年日本の輸出内訳の上位に入っている半導体などの工業関連企業は，日本の輸出額が年々上昇していることに合わせて株価も上昇しているのではないのかと考えた．