

## 研究内容

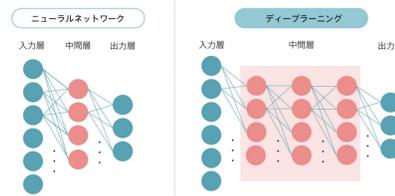
### 株価とは

企業事業の運営に必要な資金を、株を発行し売ることによって調達する。業績の良し悪しとその株式の需要によって株価が上下する。業績が良ければ株価は上昇する。一方、業績が悪くなれば株価は下落する。今回は時系列データ（株価）を機械学習させ、各手法の考察を行った。



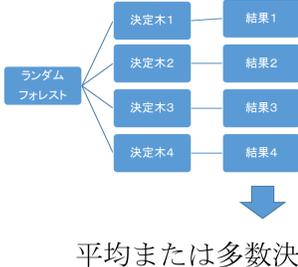
### ニューラルネットワークとディープラーニング

ニューラルネットワークとディープラーニングの大きな違いとしては中間層の差である。ディープラーニングでは中間層を多く使用するため一般的にニューラルネットワークより精度が高いとされている。



### ランダムフォレストアルゴリズム

決定木は、分類や回帰のルールをツリー状に表したものである。ランダムフォレストアルゴリズムとは、複数の決定木の多数決によって予測を決定する機械学習の手法である。また、森を大きくしても過学習が発生しないという特徴がある。



### 実験方法

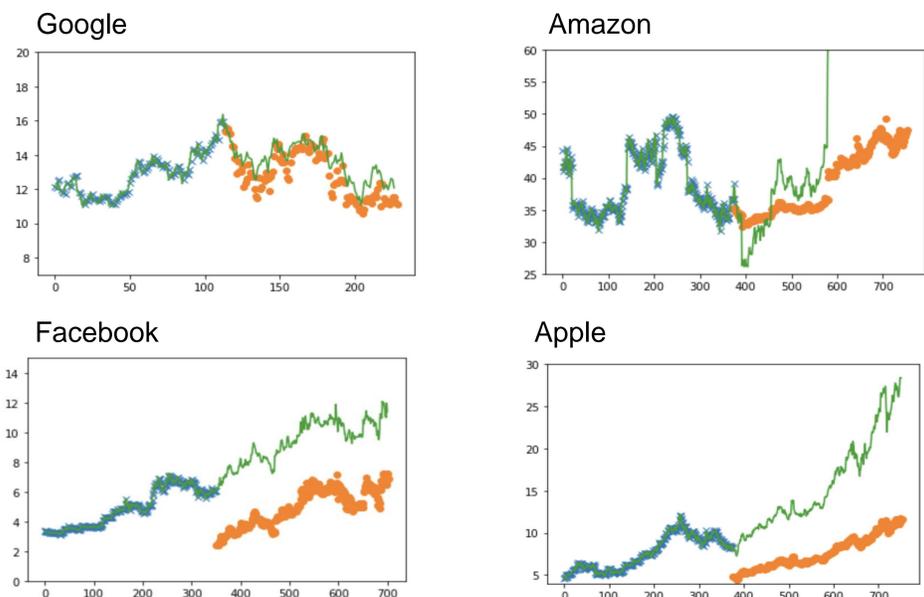
GAF(A:Google, A:Amazon, F:Facebook, A:Apple)の株式の変動をPythonを用いた機械学習の手法で株価予想を行なった。  
1,ディープラーニング  
2,ニューラルネットワーク  
3,ランダムフォレストアルゴリズム  
上記の手法の機械学習で誤差値を比較し、株価予想に適する手法について考察した。

### 仮説

ディープラーニングはニューラルネットワークを多層化し学習回数が増えるため過学習が発生し精度が下がってしまい、ランダムフォレストではそれぞれの決定木の精度を上げることが困難なため、その集合であるこの手法も精度は上がらないと考えた。上記より、ニューラルネットワークがもっとも精度が高いと予想した。

## 実験1:ニューラルネットワーク

ニューラルネットワークは、主にみずほ銀行の株取引や不動産取引などで活用されている。ニューラルネットワークでは主に株取引で使用されており今回の三つの中では一番優れていると考えた。ディープラーニングよりニューラルネットワークでは層が少ないため今回統一している60,000回という回数では、過学習を起こすことはなかった。



実際の株価と予測した値から誤差率を算出した。これは実験3のディープラーニングとも同じ計算方法である。下記の計算式で相対誤差を算出した。

$$\text{相対誤差} = (\text{実測値} - \text{理論値}) \div \text{理論値}$$

相対誤差が小さいほど予測精度が高いといえる。

| 銘柄   | Google | Amazon | Facebook | Apple |
|------|--------|--------|----------|-------|
| 相対誤差 | 8.3    | 71.4   | 55.5     | 64.2  |

以上の結果から相対誤差 49.8 となった。

## 実験2:ランダムフォレストアルゴリズム

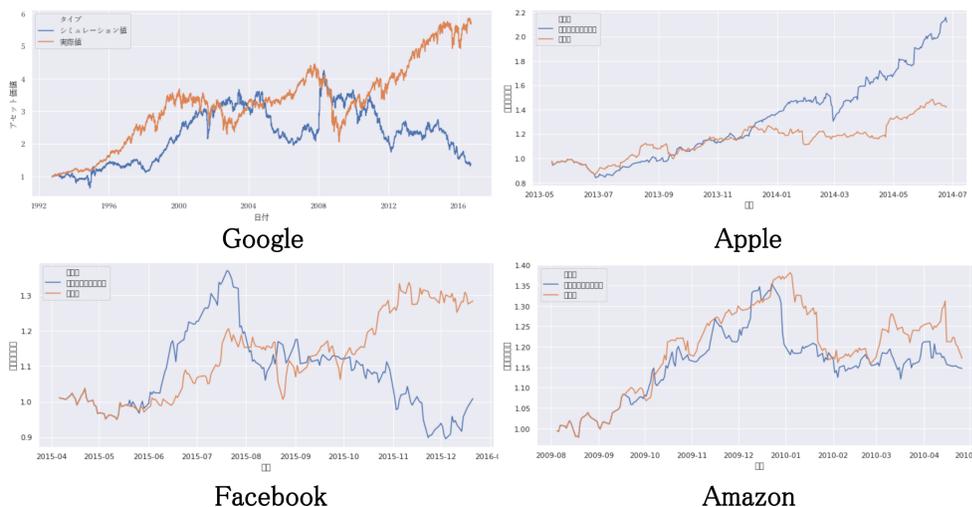
ランダムフォレストアルゴリズムは、複雑でない分類の場合、分類のプロセスを可視化することで、どの要素が結果に大きな影響を与えたのかを一目見て理解することができる。また、手書きの文字を認識したり、人体の姿勢を推定するアルゴリズムにも応用されている。

| 銘柄 | Google | Amazon | Facebook | Apple |
|----|--------|--------|----------|-------|
| 誤差 | 51.1   | 50.5   | 48.7     | 47.1  |

今回の実験では、株価が上昇、下降のどちらをするか、というプログラムを構築し検証したため(誤差)=1-(正解率)とした。結果、ランダムフォレストアルゴリズムでの平均誤差は49.4%であった。高い精度で予測が出来なかったのは、学習時に用いた変数の数値が類似していたためではないかと考える。また、ランダムフォレストアルゴリズムの性質上、それぞれの決定木の精度を上げることが難しかった今回の実験では、その集合で多数決をとるランダムフォレストの手法においても決定木の性能の限界が現れたのではないかと考えた。

## 実験3:ディープラーニング

ディープラーニングは主に物体認識や画像認識などで利用されていることが多くみられる。それは中間層を増やし学習の回数も増え精度を上げることができるからだ。しかし「株価」という日付と各項目の数値という単純さが故に、ディープラーニングの特徴である中間層の多さが影響し学習の回数60,000回では「過学習」を起こし精度が落ちてしまう。そこで、中間層などを調整し過学習しない状態を作り出し、今回の実験対象であるGAF(A)について株価予想を行なった。株価の実測値と予測値をそれぞれ比較し各銘柄の精度を相対誤差を算出した。



| 銘柄   | Google | Amazon | Facebook | Apple |
|------|--------|--------|----------|-------|
| 相対誤差 | 79.3   | 13.8   | 30.7     | 48.2  |

以上の結果よりディープラーニングによる相対誤差の平均は43.0となった。

## 考察・まとめ

| 手法   | ディープラーニング | ニューラルネットワーク | 手法 | ランダムフォレスト |
|------|-----------|-------------|----|-----------|
| 相対誤差 | 43.0      | 49.8        | 誤差 | 49.4      |

三つの実験の結果より、ディープラーニングとランダムフォレストアルゴリズムとニューラルネットワークの精度を比較したところもっとも実際の値に近かったのはディープラーニングだということが分かった。ランダムフォレストアルゴリズムでは学習に用いた変数の値が類似していたため、高い精度での予測は出来なかったと考えられる。

ニューラルネットワークでは、単純に中間層がディープラーニングよりも少ないことにより合計の学習回数が低下していたことにより精度が低下してしまったと考えられる。

以上の理由より、三つの手法の中で株価予想に最も適している手法はディープラーニングといえる。

但し、株価予測は昔のデータを学習し予測を行なっているため新商品などの予測が困難なため機械学習より株式トレーダーの方が精度が高いと言える。

### 参考文献

- [1]山口 達輝”機械学習&ディープラーニングのしくみと技術がこれ一冊でしっかりわかる教科書”株式会社アイデミー(2019)
- [2]波部 斉”ランダムフォレストの基礎と最近の動向”(2016)
- [3]大関 真之”pythonで機械学習入門 深層学習から敵対的生成ネットワークまで”オーム社(2019)
- [4]クオンツ・リサーチ株式会社”TensorFlowではじめる 株式投資のためのディープラーニング”秀和システム(2019)
- [5]川島 賢”今すぐ試したい！機械学習・深層学習(ディープラーニング)画像認識プログラミングレシピ”秀和システム(2019)