

ALife(人工生命)理論を用いた株価予測

慶應義塾大学理工学部 柳辺十武

1 はじめに

時系列データの処理に適している LSTM や、自然言語処理による感情分析などの議論が深まるにつれ、多くの人が様々な手法やアプローチで株価予測に取り組んできた。しかし必ずしもそれまでの履歴から将来が予測できるわけではなく、様々な外乱によって大きく影響を受けてしまう株価を予測することは結局のところ、不可能であると考えられることが多い。これらはそもそも機械学習という手法が莫大な情報を整理して「最適化」する力に優れているだけであるためであって、このような未来を予測するには適していないというだけである。そこで今回の提案手法ではこれらの現状を打開する ALife(人工生命)の考え方を導入して株価予測に試みてみた。

ALife とは、先述の「最適化」に優れた人工知能と比較して「新たな自然を作り出す」という特徴を持つ。ダーウィンの進化論に遡る「ありえたかもしれない生命」という考え方をもとに、「新しいものは以前のものをベースに創られる」という“隣接可能領域”という概念や、小さな相互作用が大きな群れを生む、といった考えを持つ。これはまさに過去からの履歴によって様々な相互作用を受ける株価に応用できるのではないかと考え、**銘柄を一つの生命としたモデリング**を試みる。

2 提案モデル

今回提案する手法は以下図 1 に示すものである。

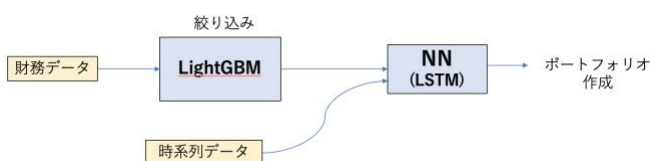


図 1 提案モデルの簡略図

まず企業の財務データからある程度の株価上昇率を LightGBM で予測し、2 月に上がるであろう銘柄を絞り込む。その後、絞り込んだ銘柄を人工生命理論を導入した NN でモデリングし、最終的な銘柄を選定、ポートフォリオを組むことにした。

2-1 LightGBM による絞り込み

LightGBM とは Microsoft が関わる Gradient Boost ライブラリの一つであり、機械学習コンペ“Kaggle”でよく話題に上がる手法である。これは決定木の弱識別器を損失が小さくなる方向に並列に並べていくというものである。LightGBM による絞り込みに際しては、重要と考えられる財務指標から学習を行うことを考えた。しかしこれら株価に影響を与えられ得る指標にも無駄なデータは含まれており、まずはこれらの指標の絞り込みから行った。指標の絞り込みを行う上で参考にしたのは以下の三点である。

- ・ LightGBM の重要度による絞り込み
- ・ Lasso 回帰による絞り込み
- ・ 統計的手法による絞り込み

以上三点による定量的な比較による指標の絞り込みを行い使用する指標を決定し、最終的な LightGBM による銘柄の絞り込みを行った。

2-2 人工生命理論を導入した NN による選定

今回の提案手法で最もポイントとなるこの NN は「自分の入力に対する相手の応答履歴から相手のモデルを学習し、これを使って次の相手の行動を予測する」という「Coupled Dynamical Recognizer; CDR」の理論をもとにする^[1]。これを実現するため、まず二つの LSTM ベースのモデルを作る(以下図 2: モデル A、モデル B)。これら二つのモデルを、モデル A は対象企業の時系列データを学習させることで、時系列によるその企業の株価変動の傾向を捉え、モデル B ではその対象企業の属する業界全体の平均株価の時系列データを学習させ、業界の株価の変動による対象企業の株価変動の傾

向を捉えさせる。ここで注意すべき点としては、与える学習データは異なるが教師データとなるものは対象企業の株価であり、共通であるという点である。

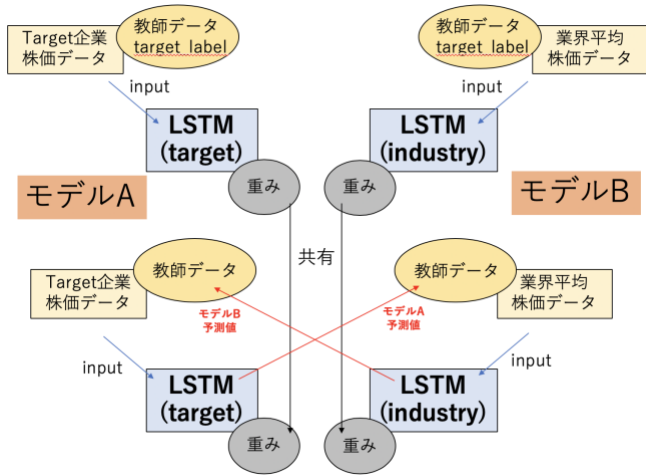


図 2 構築 CDR モデル

実際の正解データで学習を行ったモデルの重みを保存したのち、次は先ほど同様二つのモデルで学習を行う。しかし次のステップの学習はこれら対象企業のデータをもとにした(モデル A)と業界データをもとにした(モデル B)の相互作用を作り出していくことを考える。というのも先ほどの 2 つのモデルが未来の株価を予測するモデルとして成り立っていると仮定した際、お互いが相手の出力を予測し合うことで業界全体のマクロな視点と対象企業のミクロな視点を併せ持つことができると考えられる。このような相手の出力を予測し合い、動きを変えろという相互作用により、新たな生命的な動きを再現することができ、過去からの履歴が新たな自然を作り出すという ALife の理論に沿うことができる。

これら二つのモデルの出力の分析により最終的な予測株価を導き出し利益が最大化すると考えるポートフォリオを組むことを今回の提案モデルとする。

3 使用データの種類・加工

これまでに述べた理論の前提として、最低限の LightGBM による上昇可能性を持つ企業の絞り込みと株価の変動予測をするに妥当な LSTM モデルを組

むことが必要となってくる。そのためのデータ加工の工夫と過程を以下で示す。

3-1 業界からの銘柄絞り込み

モデルの指標の絞り込みを行う前に、扱う業種を限定することにした。これは業界によって特徴となる指標は異なると考え、限定した業界に適したモデルを創りあげていくことが前提条件を満たすと考えたためである。そこで今回は機械・電子機器業界に絞ることにした。これは最近話題となる twitter などでの突発的な被害によって株価に悪影響が生じる食品業界などと比べて、安定した傾向がつかめるのではないかと考えたためである。

またこの LightGBM における分析では年度通期分の財務指標値を入力とするために予測対象が決算短信発表後から任意の時期までの株価変化になってしまう。そこで今回の運用と照らし合わせ、株価変化率を決算短信発表後から年度末までの株価とし予測を行った。そして 2018 年度始めから、2018 年度末までに株価が上昇する銘柄を選定し、その中でまだ(2/8 の時点で)上昇していない銘柄が 2 月に上昇する銘柄と仮定して絞り込み作業を行うこととする。

次に LightGBM による指標の絞り込みは”2-1 LightGBM による絞り込み”で述べた三点をもとに行った。始めに取り上げた財務指標(表 1)とそれらから絞り込んだ指標(表 2)を以下に示す。

表 1 絞り込み前指標

| 使用指標名(絞り込み前) | | |
|-----------------|-------------|-----------|
| 週次株価を元にした変動係数 | 株主資本比率 | PER |
| 売上高合計(会社予想) | 企業コード | PER(会社予想) |
| EBITDA マージン | 手元流動性比率 | ROE |
| 営業活動によるキャッシュフロー | 固定比率 | 特別利益 |
| インタレストカバレッジ・レシオ | 当座比率 | 特別損失 |
| 営業 CF マージン | 有形固定資産 | 裁量的発生高 |
| フリーキャッシュフロー | 配当利回り | 株価変化率 |
| 現金同等物及び短期性有価証券 | 償却性固定資産額 | 企業価値 |
| 売上債権回転率 | 無形固定資産 | 土地 |
| 売上高研究開発費率 | 売上高合計 | EPS |
| 有形固定資産回転率 | 売上債権 | ROA |
| 企業価値(会社予想) | 企業価値/EBITDA | 配当性向 |

表2 絞り込み後の指標

| 使用指標(絞り込み後) | | |
|---------------|-------------|-------|
| EBITDA マージン | 手元流動性比率 | ROE |
| 営業 CF マージン | 株主資本比率 | ROA |
| 売上高研究開発費率 | 売上債権回転率 | 株価変化率 |
| 有形固定資産回転率 | 企業価値/EBITDA | 企業コード |
| 週次株価を元にした変動係数 | 裁量的発生高 | 配当利回り |

※使用指標は上記表 2 の指標とそれら(企業コードを除く)の前期比変化率を含む。

3-2 CDR による時系列データの取り扱い

次に CDR モデルで使用する時系列データの加工について説明する。実際の目的に沿う形で output を出すために単に株価を値とするのではなく株価の変動額をデータの形とした。初めはそのまま日毎の終値を用いていたが、株価において日毎の終値による変動額は傾向に沿わないと考えられるパターンも多く、日によって大きく異なってしまう場合があることから学習するには向いていないと考えた。そのため実際のモデルで用いたデータセットは以下図3のようなある一定範囲の変動額の平均値を一つの値とし、それらの時系列変化を入力とした。運用期間のことを考え、予測する対象変動額は 20 日後の値とし、これを 20 日後の前々日からの平均値と想定した。

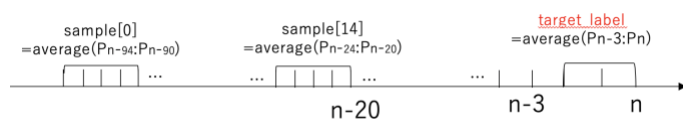


図3 データの加工模式図

また使用するデータ群は近年の企業と業界の動向を掴むために 2016 年度~2018 年度のデータのみを用いた。

4 予測結果に関する考察

4-1 銘柄絞り込み(LightGBM)

LightGBM による予測で、予測上昇率が 1.2 を上回る銘柄を以下表 3 に示す。

表3 予測上昇率>1.2 となる銘柄

| 企業名称 | 予測上昇率 | 前年変化率 | 暫定上昇率 | 差分(潜在的上昇率) |
|--------------|-------|--------|--------|------------|
| ナ・デックス | 1.261 | 1.921 | 0.7490 | 0.5124 |
| ホシザキ | 1.203 | 0.8555 | 0.8141 | 0.3893 |
| ローツェ | 1.207 | 2.736 | 0.6544 | 0.5523 |
| 愛知電機 | 1.250 | 1.693 | 0.8467 | 0.4035 |
| 寺崎電気産業 | 1.246 | 1.342 | 0.6608 | 0.5848 |
| 酒井重工業 | 1.260 | 1.853 | 0.5010 | 0.7588 |
| 象印マホービン | 1.213 | 0.9255 | 0.7945 | 0.4181 |
| 中西製作所 | 1.318 | 1.522 | 0.6697 | 0.6481 |
| 中野冷機 | 1.305 | 1.217 | 1.4140 | -0.1085 |
| 平田機工 | 1.420 | 2.113 | 0.6181 | 0.8023 |
| 野村マイクロ・サイエンス | 1.413 | 2.898 | 0.5807 | 0.8319 |

これらの分析から、前年度からの上昇率が高い企業は全体的に次年度の予測値が高くなる傾向があり、これは直感的なものに沿うと言える。また 2000~2014 年度までの値を学習データとして 2015~2017 年度をテストデータとした際に、1.4 倍以上と予測した銘柄が実際 1.1 倍以上となる確率は 76.4%、1.3 倍以上と予測して 1.1 倍以上となるのは 61.0%となった。これらから以上 11 銘柄を絞り込み結果として、NN の段階へと進めることにした。

4-2 CDR による銘柄選定

モデルとなる LSTM はそれぞれ Dropout(0.8)を持つ 2 層から成り、隠れ層はともに 128 で設定をした。企業に特化したモデルを各企業毎に創りあげていけば良いので、ある程度の過学習を許容し単純な損失をもとに学習を行った。また自分の入力に対する相手の応答履歴を学習する際に、2016~2018 年度のデータを三分割し、それぞれの段階で学習・出力を繰り返し、その出力を正解データとした。詳細を以下図 4 に示し、

この四角で覆われた行程を計 3 回繰り返した。

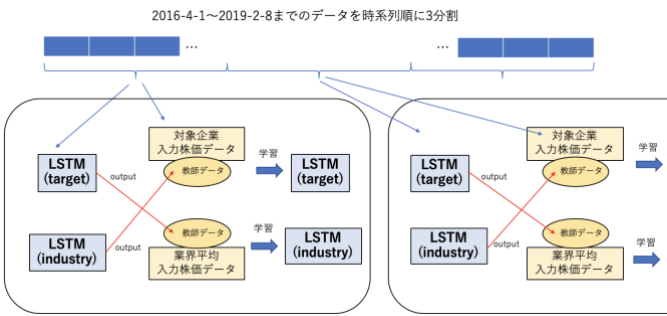


図 4 CDR 学習の概要図

この時注意すべき点としては、徐々に与えるデータの日時を最近にすることと、繰り返しの過多による予測値の収束を避けることである。

最近 10 データをテストデータとして学習を行い、それら出力と正解データを比較した際の特徴的な銘柄について以下に示す。

表 4 銘柄別予測と正解データの比較

| 中西製作所 | | | 象印マホービン | | |
|--------|--------|----------|---------|--------|----------|
| y_test | y_pred | y_pred | y_test | y_pred | y_pred |
| | target | industry | | target | industry |
| 14.27 | 1.38 | 20.37 | 177.53 | -5.05 | -15.96 |
| 14.87 | 1.13 | 22.05 | 201.67 | -5.25 | -14.84 |
| 10.80 | 5.30 | 24.83 | 215.60 | -5.12 | -13.17 |
| 32.47 | 5.22 | 23.03 | 222.67 | -6.28 | -13.89 |
| 56.27 | 2.54 | 14.58 | 193.67 | -6.56 | -15.94 |
| 67.60 | 3.15 | 16.60 | 155.60 | -6.10 | -15.72 |
| 84.67 | 3.28 | 14.84 | 136.33 | -6.27 | -17.53 |
| 97.67 | -3.02 | -1.55 | 107.20 | -6.48 | -19.42 |
| 86.93 | -3.66 | -6.72 | 75.67 | -5.40 | -20.49 |

上記表左銘柄のようにうまく特徴を捉え、変動額をモデルそれぞれが近しく予測する銘柄もあれば、全く予測を行えていない銘柄(表右)も存在した。

これら結果を踏まえて、最終的に全データを与えた時の 20 日後の予測値を二つのモデルでそれぞれの銘柄で出力させた。

そしてそれら CDR モデルによる出力と LightGBM

での結果も考慮に含め、ポートフォリオを作成した。選定銘柄とそれらの予測値について以下表 5 にまとめる。

表 5 最終選定銘柄

| 企業名称 | 予測上昇率 | 企業モデル | 業界モデル |
|---------|-----------|-------|-------|
| | /LightGBM | 予測 | 予測 |
| 愛知電機 | 1.250 | 85.65 | 31.89 |
| 中西製作所 | 1.318 | 6.139 | 28.19 |
| 中野冷機 | 1.305 | 97.44 | 103.3 |
| 平田機工 | 1.420 | -0.73 | -0.80 |
| 野村マイク | | | |
| ロ・サイエンス | 1.413 | 7.830 | 16.28 |

CDR によるモデリングが銘柄によって精度が異なるために選定においては LightGBM での分析も考慮に入れた。平田機工のみモデル予測がマイナスの値になってしまったがポートフォリオに組み入れたのはそのような理由である。

5 おわりに

本提案手法では、ALife(人工生命)に基づいた、相互作用により新たな動きを予測する CDR モデルを用いて株価を予測することを試みた。具体的には、LSTM ベースの二つのモデルで自らの入力に対する相手の応答履歴から相手の動きを予測するという手法で、銘柄の絞り込みを行った。ある程度の傾向をつかんだ未来の動きを予測できる銘柄も存在することを確認したと同時に、あくまで表面的な株価の変動からの予測でしかなく、急な変動や過去に存在しないパターンには対応ができず、銘柄によって予測精度が大きく異なってしまうことも確認した。

今後の課題としては、さらに人工生命理論に踏み込んだ応用を試みてみる。そこで、生命の**身体性**として NN に財務指標のような企業の特徴を表すデータを踏まえたり、生命の**思考・感情**として言語処理の感情分析などを導入することで深い意味での**銘柄を生命とみなしたモデリング**ができるのではないかと考える。

参考文献

[1]Takashi Ikegami and Makoto Tajii, Imitation and Cooperation in Coupled Dynamical Recognizers. Advances in Artificial Life. eds. Floreano,D.et al.,