

深層学習を用いたアンサンブルモデルによる企業価値推定モデルの提案

Model for Evaluation of Enterprise Values by Ensemble Model Using Deep Learning

田村浩一郎
Koichiro Tamura

松尾 豊 上野山勝也
Yutaka Matsuo Katuya Uenoyama

東京大学

Since 2016, in Japan, Finance Services Agency began to introduce Financial Disclosure system, which limits the selective disclosure of company's information before a statement of accounts. As a result, it became difficult for stock analysts to get contact with companies and accuracy of stock analyst's evaluation and prediction have been decreased. The models to evaluate enterprise and predict stock price which many stock analysts use mainly focus on the single company or single sector by having scrutiny investigations. However, since we can get a lot of information about all companies from the Internet now, we can evaluate companies value and predict stock price cross-sectionally and numerically. In this research, Evaluation enterprise value and stock prediction are done by some machine learning techniques, such as ensemble model and LSTM, one of the deep learning methods. From the experiments, we can tell that the proposed model can predict feature's stock price more accurately and have the possibility to combine other prediction methods to improve the accuracy.

1. 背景

近年、金融庁は、中長期的な投資を推奨することを目的に、企業が未公表の重要情報を選択的に伝えることを禁じる「フェアディスクロージャー(FD)規制」の導入を進めている[1]。決算前の企業に対する証券アナリストの取材活動が規制され始めると、証券アナリストの予測が当たらなくなってしまった。その結果、金融庁の思惑に反して、決算発表後の短期的な売買が増えている[1]。証券アナリストが用いている既存の企業価値推定モデルは、調査対象となる単一業界や単一企業へのインタビュー調査から、演繹的に価値推定を行う手法が主流であったが、多くの企業情報が Web 上に公開されている現在、多数の企業を対象に横断的・解析的に企業価値を推定することが可能となっている。本研究では特に、近年飛躍的に技術が向上している機械学習技術に注目し、帰納的に企業価値を推定する手法に関して研究する。

2. 関連研究と本研究の学術的知見

既存の企業価値推定モデルには主として、資産の価値から分析するコストアプローチ、他社との比較に観点を置いたマーケットアプローチ、収益から分析するインカムアプローチといったファンダメンタル分析がある(以下、経済学的ファンダメンタルアプローチ)。経済学的ファンダメンタルアプローチは、将来キャッシュフローを始めとした入力変数や、手法の組み合わせ方に大きく依存しており、その入力変数の値や手法の組み合わせ方は、既存概念や経験に基づき演繹と帰納の繰り返して属人的に決定されている。証券アナリストは、インタビューなどを通して対象企業や業界を「深く」分析し、解像度の高い情報を元に企業価値推定を行っていたが、証券アナリストの事前調査の規制の動きにより、事前情報の入手が難しくなり、予測が当たらなくなってしまったと考えられる。

既存の企業価値推定モデルが、事前情報なしには適切な予測ができない理由として、1. 帰納的推論による予測の最適化が効果的に行われていない 2. 複数の手法が孤立している という二つの潜在的な問題が挙げられる。人間による帰納的推論による予測手法には、物理的制約によるデータサンプル数の限界と、認知限界があるため、効果的に帰納的推論による予測が行われていない。また、上述したような複数のファンダメンタル分析の他にも、株価予測手法として、市場の指標によるテクニカル分析など複数提案されており、株価予測手法に関しては機械学習を用いた研究も多い。しかし、筆者が知りうる限り複数の異なる手法を組み合わせるモデルが体系立てられておらず、複数の既存手法が孤立してしまっており、汎化性能が高められていない。

本研究では、日本の全上場企業に対して機械学習を用いた横断的・解析的な手法を試みることで、これらの既存手法の問題を解決し、より優れた企業価値推定を行うことを目的とする。

本研究で提案する企業価値推定モデルは、既存の企業価値推定手法と異なり、機械学習を用いた帰納的手法であるという点で経済学的分野において新規性がある。また、既存の機械学習を用いた株価予測手法の関連研究とは、予測対象が指数でなく企業の株価であり、企業活動に関するファンダメンタルな因子に着目し横断的な分析である点、及び複数の手法を融合させる手法を考案している点で本質的に異なる。

3. 関連研究と本研究の学術的知見

(1) 提案モデル概要

提案モデルは図.1 のように大きく 3 つの要素で構成される。財務情報を入力素性とした機械学習によるファンダメンタル分析のモデル(以下、提案モデルファンダメンタルアプローチ)、Long short term memory[2](以下 LSTM)を用いて企業間の時系列相関を学習したテクニカル分析のモデル(以下、提案モデルのテクニカルアプローチ)、そしてそれらを融合させるアンサンブルモデル(以下、提案モデルの最終層)である。提案モデルのフ

連絡先: 田村浩一郎 東京大学 koichiro11@weblab.t.u-tokyo.ac.jp

ファンダメンタルアプローチおよびテクニカルアプローチ(以下、両者を合わせて提案モデルの一層目と呼ぶ)は、機械学習による十分なデータサンプルを用いた帰納的推論による推定を可能にするだけでなく、End-to-Endのアルゴリズムであることによってファンダメンタル因子の中の市場の価格決定要因や、市場間の時系列相関を抽出することが期待できる。また、提案モデルのファンダメンタルアプローチによる推定値と提案モデルのテクニカルアプローチの推定値を非線形モデルである XGBoost[3]の入力素性とする事でアンサンブルモデル[4]を構築し、複数の手法を非線形に組み合わせることを可能にし、汎化性能を高めることが期待できる。

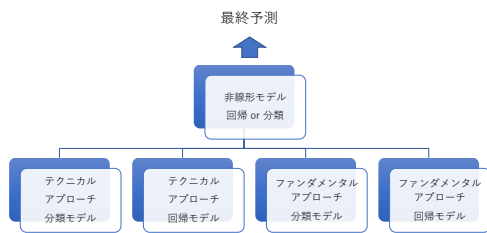


図.1 提案モデルの概要

(2) 提案モデルのファンダメンタルアプローチ

提案モデルのファンダメンタルアプローチの概要は図.2の通り。

XBRLファイル取得、解析することで企業活動情報をデータベース化するシステムを構築した。そして、取得したデータセットに対して XGBoost を用いて特徴量ごとの重要度を、情報利得を基に計算し、それに基づいて特徴量抽出を行う。特徴量抽出によって次元削減された特徴量を用いて複数の学習器を学習させ、各学習器の予測を用いて XGBoost を学習・予測をするアンサンブルモデルを構築する。

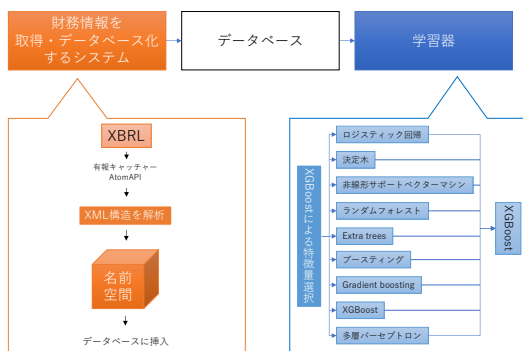


図.2 提案モデルのファンダメンタルアプローチの概要

(3) 提案モデルのテクニカルアプローチ

提案モデルのファンダメンタルアプローチは再帰的深層学習の一つである LSTM を用いることで、勾配が消滅して対象データの長期的な特徴量を捉えることができないという勾配消滅問題の解決を図り、より長期的な企業間の時系列相関を学習することを期待する。

4. 関連研究と本研究の学術的知見

証券アナリストの予測と同様に、予測対象は各企業の投資判断(買いか売るか)及び目標株価とし、各企業の決算短信発表日から営業 60&240 日で検証する。また、特定の期間に過学習したモデルでないことを示すために、検証期間は 2013 年第 1 四半期から 2014 年第 4 四半期の計 8 期間とする。提案モデルのファンダメンタルアプローチでは、上場企業約 3000 企業の 5 年分の決算データ、総計約 60000 サンプルを学習データセットとする、提案モデルのテクニカルアプローチでは、日経 225 平均株価の構成銘柄を対象に 5 年分の週足データ、総計約 56000 サンプルを学習データセットとして用いる。

まず、提案モデルのファンダメンタルアプローチと既存の経済学的ファンダメンタルアプローチの精度比較を行う。精度比較の結果は以下のとおり。

(1) 分類予測の場合

経済学的ファンダメンタルアプローチの平均正解率=0.5019 < 提案モデルのファンダメンタルアプローチの平均正解率=0.6355, 16 検証パターン中 14 検証パターンで提案モデルが上回る

(2) 回帰予測の場合

経済学的ファンダメンタルアプローチの誤差率平均=29.51% < 提案モデルのファンダメンタルアプローチの正解率=14.41%, 16 検証パターン中 16 検証パターンで提案モデルが上回る

次に、他企業のテクニカル因子を入力素性に加えない LSTM のモデル(以下既存のテクニカルアプローチ)と、提案モデルのテクニカルアプローチを比較検証する。精度の比較検証の結果は以下のとおり。

(3) 分類予測の場合

既存のテクニカルアプローチの平均正解率=0.4905 < 提案モデルのテクニカルアプローチの平均正解率=0.5272, 8 検証パターン中 6 検証パターンで提案モデルが上回る

(4) 回帰予測の場合

既存のテクニカルアプローチの誤差率平均=8.22% < 提案モデルのテクニカルアプローチの誤差率平均=7.85%, 8 検証パターン中 6 検証パターンで提案モデルが上回る

そして、提案モデルの一層目と提案モデルの最終層の精度を比較検証した。結果は以下の通り。

(5) 分類予測の場合

提案モデルの一層目の平均正解率=0.5779 < 提案モデルの最終層の平均正解率=0.6395, 16 検証パターン中 8 検証パターンで提案モデルの最終層が上回る

(6) 回帰予測の場合

提案モデルの一層目の誤差率平均=14.95% < 提案モデルの最終層の誤差率平均=11.73%, 16 検証パターン中 14 検証パターンで提案モデルの最終層が上回る

さらにこの検証の過程で、テクニカルアプローチとファンダメンタルアプローチが予測問題・期間によって予測に対する優位性が異なることが示された。また、提案モデルの最終層と市場との比較(分類予測なら市場のトレンド分布, 回帰予測なら現在値を予測としたもの。以下真のベースラインと呼ぶ)を行うと、特に回帰予測においては真のベースラインを上回ることが検証された。

5. 結論

検証実験によって、提案モデルは、データサンプル数を増やした帰納的推論による予測を可能にし、また、学習過程で市場の価格決定要因や時系列相関を抽出することによって、既存手法よりも優れた予測を行えることが示された。そして、アンサンブルモデルによって異なるアプローチを融合させた提案モデルが、より汎化性能の高いモデルとなることを示し、特に回帰予測において実用的な予測精度であることを示した。この結果は、提案モデルによって、今まで独立に提案されていた複数の既存手法を予測問題に応じて組み合わせることが可能であり、さらに優れた予測を行えるようになる可能性を示唆している。

6. 謝辞

本研究は JSPS 科研費 JP25700032, JP15H05327, JP16H06562 の助成を受けたものです。

参考文献

- 1) 川崎健. アナリスト不在の不幸 強まる規制、荒れる株価.
日本経済新聞電子版, 11 2016.
- 2) Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- 3) Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *arXiv preprint arXiv:1603.02754*, 2016.
- 4) Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pp. 1–15. Springer, 2000.