

令和3年度

博士論文

金融市場の予測モデルにおける
ヒトの認知限界と深層学習に関する研究

令和4年3月

指導教員 松尾豊 教授

東京大学工学系研究科

技術経営戦略学専攻

37-197171 田村浩一郎

概要

金融市場における価格形成は多様な因子が相互に影響しあう複雑な現象の結果であり、予測の可能性やモデリングの手法に関して、学術的にも実務的にも関心を集めてきた。金融市場は複雑であるが故に、既存の金融経済学では、複雑な事象をヒトが理解できるように単純化するため、いくつかの前提仮説を導入し、特定の切り口で理論化を進めてきた。

効率的市場仮説 [1] は、「全ての情報がすでに株価に織り込まれている」という金融経済学において重要な仮説である。実際には、アノマリー [2] の存在や、投資家は限定合理性 [3] の中で意思決定を行うために、市場全体の合理性が成り立たないという指摘がなされている。効率的市場仮説が常に成立することはなければ、金融市場を予測することで、市場平均を上回る期待収益 α を追求することも可能になる。

しかし、金融市場の予測は本来的に難しい。「効率的市場仮説が常に成立することはない」としても、基本的には市場は効率的で、十分に多くの投資家は瞬時に株価やニュースなどの情報を得ることができる。また、金融市場は非定常的である。例えばアノマリーを発見して予測モデルを構築し、一時的に収益を得ることはできても、安定的な収益を得られるモデルを構築することは困難である。

本研究では、複雑な金融市場をヒトが理解できるよう単純化するために導入された、既存の金融理論の前提仮説や特定の切り口に対して、予測可能性を見出す。ヒトは数千、数万ある金融商品の価格を瞬時に認識することはできず、常時情報処理を行うこともできない。ヒトの認知処理限界を理由に単純化されていた既存の金融理論の手法を拡張することができれば、より精緻な予測が可能になると仮説を立てることができる。

深層学習は、従来の特徴量抽出器・分類器よりも優れた性能を発揮し、特に学習過程で隠れ層において入力特徴空間に依存しない特徴表現を獲得できるアルゴリ

ズムである。ヒトの認知処理能力では取り扱うことのできないほどの膨大な時空間的データを処理できる可能性や、帰納的に重要な素性や関係性を抽出しモデリングすることができる可能性を持つ。

既存の金融理論の中でヒトの認知処理限界を理由に単純化されていた手法に対して深層学習を用いることによって、人の認知処理限界に依存しない金融理論を構築できる可能性がある。

今までヒトの認知処理限界が背景にあったモデルを、深層学習を用いたアルゴリズムで拡張するためには、既存のモデルにおいてヒトの認知処理限界がボトルネックになっている要素の仮説を構築する必要がある。深層学習は前節で述べたとおり優れたアルゴリズムであるが、データにノイズも多く含まれ、非定常性がある金融市場においては、予測モデル全体を深層学習で置き換えただけでは上手くいかない可能性がある。実際、機械学習を用いたモデルよりも、伝統的な統計的手法を用いたモデルの方が優れている可能性を示唆する主張も存在する [4]。

そこで本研究は、データ構造の画一性が低く、情報の非対称性が相対的に大きい ii) 金融資産、 iii) 投資家を分析することで金融市場を予測する手法を対象にする。

まず、金融資産の特徴や性質に着目して株主価値を予測する際、ヒトでは1つないしは数企業の財務モデリングを行う事が限界であった手法に対して、深層学習を用いて複数の企業の財務・業績情報を同時に取り扱う手法に拡張する。既存のテクニカル分析の手法を用いるだけでなく、財務・業績情報を用いて市場における価値に着目したファンダメンタルズ分析を機械学習を用いて行い、アンサンブル学習を用いた複合的な株主価値推定モデルを提案した。XBRL ファイルを利用し、全ての上場企業の企業情報を取得するシステムを用いて企業情報を一元化した。推定対象を標準化された株主価値にすることで、機械学習によるファンダメンタルズ分析を可能にした。機械学習によって、約 60000 サンプルのデータセットから帰納的な推定を行ったこと、重要な特徴量を 1000 項目ほどの企業情報の特徴量から抽出することによって、人間の処理できるデータサンプル数の限界及び人間の認知限界の問題を解消し、より精度の高い株主価値推定が行えることを示した。

次に、ポートフォリオのリスクを予測する際、ヒトではポートフォリオを構成す

る全ての企業のビジネスモデルや事業内容とそのリスクの関係を知らずに限界があったのに対して、深層学習を用いて企業の事業内容のベクトル表現を獲得し、ビジネスモデルや事業内容を反映した共分散行列の予測手法に拡張する。企業のビジネスモデルや事業内容のベクトル表現 \mathbf{b}_n を用いる提案手法が、将来の共分散行列をより精緻に予測できることを示した。また、提案手法を用いて予測される共分散行列を用いて構築されたポートフォリオは、ビジネスモデルや事業内容を考慮しない場合のポートフォリオよりも優れたパフォーマンスを出すことを示した。さらに、提案手法を用いて導出されるベクトル表現 \mathbf{b}_n を用いることで、既存の業界区分やセグメント情報では適切に表現しきれない企業情報を抽出できることや、ビジネスモデルや事業内容の性質に応じた期待収益率やリスクの分析可能性が高まっていることを示唆した。

そして、投資家の心理や行動に着目して価格変動を予測する際、ヒトでは社会ネットワークの構造や自ら情報発信をしない投資家を分析することには限界があったのに対して、Deep Learning を用いてソーシャルメディア内のネットワーク構造と、直接観測することが難しい投資家の内部状態から金融市場への作用をモデル化した予測手法に拡張する。ソーシャルメディアは、金融市場において投資家が何を認知し、どのような感情を抱いているかを獲得することができるセンサとして機能しているだけでなく、金融市場の価格変動に働きかけるアクチュエータとしても機能している可能性がある。直接観測することができず、時間を通して相互に複雑に関連している、ソーシャルメディアにおける影響関係から金融市場への作用を、RNN を用いてモデル化する手法を提案した。実験を通して、ソーシャルメディアにおける影響関係から金融市場への作用をモデル化する提案手法が、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻にかつ安定して予測できることを示した。また、提案手法のモデルは、ソーシャルメディアの影響関係から金融市場への作用の分析可能性を高めた。分析の中で、ソーシャルメディアのネットワークが直接接続されていなくとも、時間を通して影響関係になる可能性があり、さらに金融資産・市場の性質によってソーシャルメディアから金融市場への影響度が異なることを示した。

目次

第1章 序論	13
1.1 背景	13
1.2 研究目的	16
1.3 本論文の構成	18
第2章 関連研究	19
2.1 金融市場の予測モデルの研究の位置づけ	19
2.1.1 金融市場の予測モデルで取り扱う情報と予測の目的	19
2.1.2 効率的市場仮説と予測モデル	21
2.2 過去の株価やニュース情報を用いて市場の状態に着目した予測モデルの研究	24
2.2.1 市場の状態に着目した時系列予測モデルの研究	25
2.2.2 モデルの入力として用いる情報ソースを拡張する研究	25
2.2.3 取り扱う金融商品・予測対象を拡張する研究	26
2.3 金融資産が持つ特徴や性質に着目した予測モデルの研究	27
2.3.1 企業価値の推定と予測	27
2.3.2 経済学的ファンダメンタルズアプローチ	29
2.3.3 経済学的ファンダメンタルズアプローチの課題と研究の位置づけ	32
2.4 ソーシャルメディアを用いて投資家に着目する予測モデルの研究	33
2.4.1 行動経済学とソーシャルメディアによる予測モデルの可能性の拡張	33
2.4.2 仮想通貨の登場と投資家心理を考慮することの重要性	35
2.4.3 ソーシャルセンサの研究	38

2.4.4	information diffusion の研究	41
2.4.5	投資家の影響関係を表す研究	47
第3章	深層学習を用いたアンサンブルモデルによる株主価値推定モデルの提案	48
3.1	はじめに	48
3.2	提案手法	51
3.2.1	概要	52
3.2.2	企業情報の取得	53
3.2.3	テクニカル分析のモデル	54
3.2.4	ファンダメンタルズ分析のモデル	56
3.3	実験	56
3.3.1	実験目的	56
3.3.2	実験する手法	57
3.3.3	検証実験の設定	60
3.3.4	実験結果	65
3.4	実験に対する考察	70
3.5	おわりに	74
第4章	企業の事業内容に着目した共分散行列による深層学習ポートフォリオ最適化	76
4.1	はじめに	76
4.2	提案手法	79
4.2.1	決算短信及び有価証券報告書のデータベース化とテキストの取得	79
4.2.2	ニューラルネットワークを用いた事業内容のベクトル表現の獲得	81
4.2.3	事業内容のベクトル表現を用いたポートフォリオの共分散行列の予測	82
4.2.4	ポートフォリオの構築	84

4.3	実験	84
4.3.1	実験設定	84
4.3.2	比較モデル	86
4.3.3	実験結果	89
4.4	分析	92
4.4.1	企業の事業内容のベクトル表現の分析	92
4.4.2	企業のビジネスモデル・事業内容と株価パフォーマンスの 関係の分析	97
4.5	終わりに	101
第5章	ソーシャルメディアにおける影響関係から金融市場に対する作用のモ デル化と分析	102
5.1	はじめに	102
5.2	提案手法	105
5.2.1	ソーシャルメディアの影響関係から金融市場への作用の定式化	105
5.2.2	end-to-endのネットワークを用いたモデル化	106
5.2.3	影響関係行列 E による影響関係の表現と抽出	106
5.2.4	認知・心理の時系列内部状態の獲得	109
5.3	実験	111
5.3.1	実験設定	111
5.3.2	比較モデル	113
5.3.3	実験結果	116
5.4	分析	118
5.4.1	ソーシャルメディアの社会ネットワークにおける影響関係の 抽出と分析	118
5.4.2	金融資産の性質に対するソーシャルメディア特徴量の予測に 対する重要度の分析	120
5.5	終わりに	124

第 6 章 考察	126
6.1 3章から5章までのまとめ	127
6.2 本研究の貢献	128
6.3 本研究の課題と限界	130
6.3.1 時系列普遍性の課題と限界	130
6.3.2 解釈・説明性とのトレードオフの存在	132
6.4 新たな問いと展望	132
6.4.1 深層学習を用いたヒトの認知処理限界の解消と金融理論の拡張可能性はどれだけあるか	133
6.4.2 深層学習によって金融市場は効率的になるのか	134
第 7 章 結論	136
参考文献	139
謝辞	152
発表文献	156

目 次

1.1	既存の金融理論に基づく金融市場の予測モデルとヒトの認知処理限界がボトルネックになっている要素の整理	17
2.1	金融市場の予測モデルの研究の位置づけ	21
2.2	過去の株価やニュース情報を用いて市場の状態に着目した予測モデルの研究の方向性の整理	24
2.3	金融資産が持つ特徴や性質に着目した予測モデルの研究の方向性の整理	33
2.4	ソーシャルメディアを用いて投資家に着目する予測モデルの研究の方向性の整理	34
2.5	2017年における金, ビットコイン, 日経平均指数の価格推移の関係. 金と日経平均には逆相関が見られるが, ビットコインと日経平均にはそのような関係はない.	37
2.6	information diffusion の 4 分類	42
3.1	深層学習を用いたアンサンブルモデルによる株主価値推定モデル	51
3.2	分類予測におけるファンダメンタルズ分析の重要度の割合	70
3.3	回帰予測におけるファンダメンタルズ分析の重要度の割合	71
4.1	提案する企業の事業内容に着目した共分散行列の獲得手法の概要	80
4.2	各手法によって構築されたポートフォリオの収益率の推移	91
4.3	日本株式市場における東証 33 業種分類のヒストグラム. 「情報・通信」と「サービス業」の数が多く, 実際に行っている事業内容を捉えきれていない可能性がある.	94

4.4	「情報・通信」に属する企業の事業内容を表すベクトル表現の3次元空間へのマッピングの結果	95
4.5	「サービス業」に属する企業の事業内容を表すベクトル表現の3次元空間へのマッピングの結果	96
4.6	LSI-BCM のトピック分析によって導出された事業内容を表すベクトル表現の3次元空間へのマッピングの結果	98
4.7	Deep-BCM を用いて導出された事業内容を表すベクトル表現の3次元空間へのマッピングの結果	99
4.8	ビジネスモデル・事業内容ごとの株価パフォーマンスの比較. SaaS 企業は相対的にパフォーマンスが高く, COVID-19 の感染拡大後も急速な回復をしている. 顧客チャネルやサプライチェーンを国内外に構築するグローバル企業は, COVID-19 後のパフォーマンスが相対的に低迷している.	100
5.1	end-to-end のネットワークを用いたモデルの概要	105
5.2	ビットコインの3日後のボラティリティ予測の検証実験で獲得された影響関係を表す行列 E において, サンプリングされた10アカウントの影響関係のヒートマップ. 値が大きいほど影響が大きいことを示している.	119
5.3	事業ごとの W_{out} の分析結果. 金融市場における価格変動を表す変数 h_t への重みに対して, ソーシャルメディアにおける事象を表す変数 s_t への重みの差分を表している. 値が大きいほどソーシャルメディアが予測に寄与する比重が大きい. バイオ事業やソーシャルゲーム事業において, よりソーシャルメディアにおける事象を表す変数 s_t が予測に寄与していることがわかる.	122
5.4	市場ごとの W_{out} の分析結果. 金融市場における価格変動を表す変数 h_t への重みに対して, ソーシャルメディアにおける事象を表す変数 s_t への重みの差分を表している. ビットコイン市場やマザーズ市場において, よりソーシャルメディアにおける事象を表す変数 s_t が予測に寄与していることがわかる.	123

表 目 次

3.1	採用するテクニカル指標一覧	62
3.2	採用するファンダメンタルズ指標一覧	64
3.3	分類予測における正解率の比較 (%)	65
3.4	回帰予測における正解率の比較 (%)	66
3.5	2014 年第 4 四半期・営業 240 日平均の分類予測におけるファンダメンタルズ指標重要度	68
3.6	2014 年第 4 四半期・営業 240 日平均の回帰予測におけるファンダメンタルズ指標重要度	69
4.1	2019 年 1 月 1 日から 2019 年 12 月 31 日までの共分散行列の予測における提案手法の RMSE 比較	90
4.2	2020 年 1 月 1 日から 2020 年 12 月 31 日までの共分散行列の予測における提案手法の RMSE 比較	90
4.3	各手法のポートフォリオのパフォーマンス評価	91
4.4	期待収益率が既知の場合の各手法のポートフォリオのパフォーマンス評価	92
5.1	提案手法と比較手法のデータ利用・構造の違い	113
5.2	ビットコインの将来のボラティリティ予測における提案手法の RMSE 比較	116
5.3	イーサリアムの将来のボラティリティ予測における提案手法の RMSE 比較	116
5.4	リップルの将来のボラティリティ予測における提案手法の RMSE 比較	116

第1章 序論

1.1 背景

金融市場における価格形成は多様な因子が相互に影響しあう複雑な現象の結果であり、予測の可能性やモデリングの手法に関して、学術的にも実務的にも関心を集めてきた。金融市場は複雑であるが故に、既存の金融経済学では、複雑な事象をヒトが理解できるように単純化するため、いくつかの前提仮説を導入し、特定の切り口で理論化を進めてきた。

効率的市場仮説 [1] は、「全ての情報がすでに株価に織り込まれている」という金融経済学において重要な仮説である。効率的市場仮説は、投資家が常に情報を瞬時的に「認知」し、正しい「判断」を行うという強力な前提を持つ。合理的期待形成仮説は、合理的であろうとする市場参加者は全体として、現時点での価格との比較で裁定取引を行いリターンを得ようとするため、近似的に市場は常に完全に情報的に効率的であるとする仮説である。個人単位ではノイズトレーダー [5] のような非合理的な行動をとる市場参加者が存在していたとしても、全体としては近似的に極めて合理的であるとすることによって、効率的市場仮説をサポートする。

効率的市場仮説や合理的期待形成仮説に基づき、ランダムウォーク理論 [6] が生まれた。ランダムウォーク理論は、株価の動きは過去の株価と独立で不規則的であり、予測をすることができないとする理論である。ランダムウォーク理論に基づくと、株価の予測そのものの可能性を否定する理論であるが、一方で、確率的には一定範囲内で価格形成がなされるのであれば、ある瞬間の静的な均衡状態をモデル化して分析可能になると考えられる。ランダムウォークする株価は正規分布に従うため、株式投資のリスクはリターンの分散として捉える事ができる。

リスクはリターンの分散という前提に基づくと、ポートフォリオのリスクも共分散を用いてモデル化する事ができる。CAPM理論 [7] は、市場ポートフォリオを導入し β を用いて単純化すれば、任意の金融資産の期待リターンを計算できるとする理論で、CAPM理論は基本的な資産価格モデルの一つであり、実用でも広く浸透している。[8] はCAPM理論のモデルとしての当てはまりの悪さを指摘し、より当てはまりの良いファーマ-フレンチの3ファクターモデルを提案した。提案者の1人である Fama はノーベル賞を受賞している。

金融経済学は、金融商品の価格形成がなされる市場や、投資家の投資行動、企業の財務状況や資本構成を分析対象とする、経済学の分野の1つと説明される。全ての情報を取り扱うことは難しいため、市場に着目するテクニカルアプローチや、企業の財務状況や資本構成に着目するファンダメンタルズアプローチ、投資家の投資行動に着目する行動経済学的アプローチなど、特定の切り口で分析を行い、予測モデルが構築されている。

効率的市場仮説が常に成立する場合、過去の情報から将来の金融市場を予測することはできないが、実際には、モメンタム効果や、小型株効果、バリューストック効果といったアノマリー [2] の存在や、投資家は限定合理性 [3] の中で意思決定を行うために、合理的市場形成仮説における市場全体の合理性が成り立たないという指摘がなされている。効率的市場仮説が常に成立することはなければ、金融市場を予測することで、市場平均を上回る期待収益 α を追求することも可能になる。 α の追求は実務家にとっても重要な課題だ。

しかし、金融市場の予測は本来的に難しい。「効率的市場仮説が常に成立することはない」としても、基本的には市場は効率的で、十分に多くの投資家は瞬時に株価やニュースなどの情報を得ることができる。また、金融市場は非定常的である。例えばアノマリーを発見して予測モデルを構築し、一時的に収益を得ることはできても、安定的な収益を得られるモデルを構築することは困難である。

そもそも金融市場の予測とは何か、というのは難しい問いである。[9]は、「予測かどうかは、主体がもつモデル及びその社会的な合意に依存する」と述べている。金融市場の予測とは、投資家がもつモデルと現時点での社会的な合意との差分を認識することであり、金融市場の予測モデルとは、投資家がもつモデルと現時点での社会的な合意との認知差分を理論化したものであると考えられる。

本研究では、複雑な金融市場をヒトが理解できるよう単純化するために導入された、既存の金融理論の前提仮説や特定の切り口に対して、予測可能性を見出す。ヒトは数千、数万ある金融商品の価格を瞬時に認識することはできず、常時情報処理を行うこともできない。ヒトの認知処理限界を理由に単純化されていた既存の金融理論の手法を拡張することができれば、より精緻な予測が可能になると仮説を立てることができる。

深層学習は、従来の特徴量抽出器・分類器よりも優れた性能を発揮し、特に学習過程で隠れ層において入力特徴空間に依存しない特徴表現を獲得できるアルゴリズムである。ヒトの認知処理能力では取り扱うことのできないほどの膨大な時空間的データを処理できる可能性や、帰納的に重要な素性や関係性を抽出しモデリングすることができる可能性を持つ。深層学習は、複雑な問題を複雑なまま解くことのできる潜在可能性を持っている。

既存の金融理論の中でヒトの認知処理限界を理由に単純化されていた手法に対して深層学習を用いることによって、人の認知処理限界に依存しない金融理論を構築できる可能性がある。また、ヒトの知見を複数再現することができれば、複数のモデルを用いた取引戦略を構築することができ、市場の非定常性にも一定対処できる可能性がある。

1.2 研究目的

本研究の大目的を以下とする.

大目的

既存の金融理論に基づく金融市場の予測手法に対して, 今までヒトの認知処理限界が背景にあったモデルを, 深層学習を用いたアルゴリズムを組み込み拡張すること

今までヒトの認知処理限界が背景にあったモデルを, 深層学習を用いたアルゴリズムで拡張するためには, 既存のモデルにおいてヒトの認知処理限界がボトルネックになっている要素の仮説を構築する必要がある. 深層学習は前節で述べたとおり優れたアルゴリズムであるが, データにノイズも多く含まれ, 非定常性がある金融市場においては, 予測モデル全体を深層学習でただ置き換えただけでは上手くいかない可能性がある. 実際, 機械学習を用いたモデルよりも, 伝統的な統計的手法を用いたモデルの方が優れている可能性を示唆する主張も存在する [4].

金融市場の予測モデルでは, 分析する対象は i) 市場, ii) 金融資産, iii) 投資家の3つに分類できる. このうち, 図 1.1 で整理されるように, データ構造の画一性が低く, 情報の非対称性が相対的に大きい ii) 金融資産, iii) 投資家を対象にする分析は, ヒトの認知処理限界による制約が大きいと考えることができる.

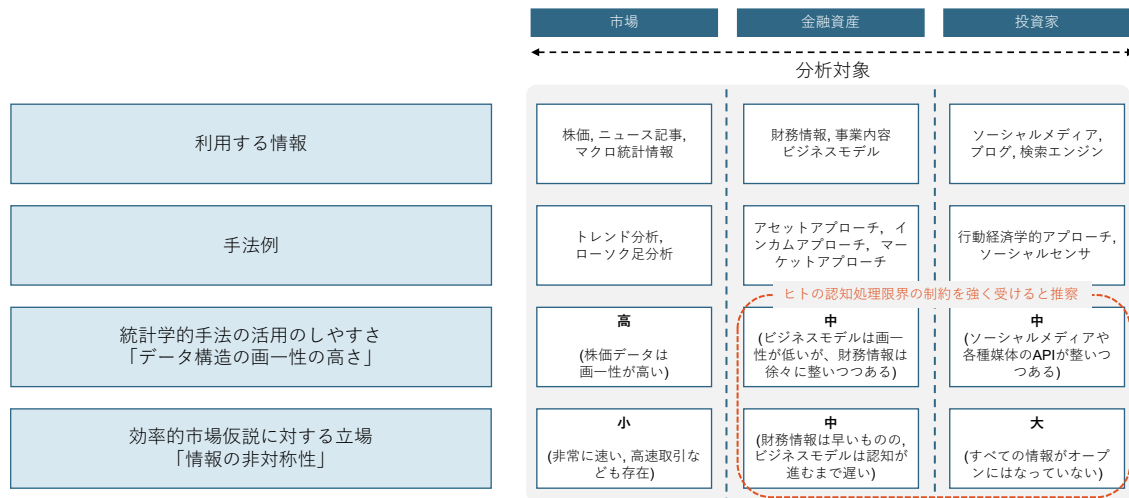


図 1.1: 既存の金融理論に基づく金融市場の予測モデルとヒトの認知処理限界がボトルネックになっている要素の整理

そこで本研究は、

目的 1 金融資産の特徴や性質に着目して株主価値を予測する際、ヒトでは1つないしは数企業の財務モデリングを行う事が限界であった手法に対して、深層学習を用いて複数の企業の財務・業績情報を同時に取り扱う手法に拡張し、その有用性を検証すること。

目的 2 ポートフォリオのリスクを予測する際、ヒトではポートフォリオを構成する全ての企業のビジネスモデルや事業内容とそのリスクの関係を知らずに限界があったのに対して、深層学習を用いて企業の事業内容のベクトル表現を獲得し、ビジネスモデルや事業内容を反映した共分散行列の予測手法に拡張し、その有用性を検証すること。

目的 3 投資家の心理や行動に着目して価格変動を予測する際、ヒトでは社会ネットワークの構造や自ら情報発信をしない投資家を分析することには限界があったのに対して、深層学習を用いてソーシャルメディア内のネットワーク構造と、直接観測することが難しい投資家の内部状態から金融市場への作用をモデル化した予測手法に拡張し、その有用性を検証すること。

を目的とする.

1.3 本論文の構成

本論文の構成について述べる.

2章では, 本論文で取り組む3つの研究に関して関連研究を整理し, それぞれの研究の学術的見地を明確にする. 3章では, 目的**1**のために, 深層学習を用いたアンサンブルモデルによる株主価値推定モデルの提案を行う. 4章では, 目的**2**のために, 企業の事業内容に着目した共分散行列による深層学習ポートフォリオ最適化の手法を提案する. 5章では, 目的**3**のために, ソーシャルメディアにおける影響関係から金融市場に対する作用のモデル化と分析を行う. 6章では3~6章の研究のまとめと本研究の貢献及び課題を述べ, 最後に7章で結論を述べる.

第2章 関連研究

本章では関連研究について整理し，本研究の学術的見地を明確にする．はじめに，金融市場の予測モデルの研究を，予測の目的や取り扱う情報を軸に整理し，金融市場における予測のタスクの解像度を高める．次に，市場の状態に着目した予測モデル，金融資産の性質に着目した予測モデル，投資家に着目した予測モデルのそれぞれに対して，既存研究を整理し述べる．

2.1 金融市場の予測モデルの研究の位置づけ

2.1.1 金融市場の予測モデルで取り扱う情報と予測の目的

金融市場における価格形成は多様な因子が相互に影響しあう複雑な現象の結果であり，予測の可能性やモデリングの手法に関して，学術的にも実務的にも関心を集めてきた．

金融市場の予測には，過去の株価情報からスーパーの駐車場の衛星写真まで，様々な情報が活用されている．金融市場の予測モデルは，着目する情報を軸に以下の3つの予測モデルに分類することができる．

1. 市場の状態に着目した予測モデル
2. 金融資産の性質に着目した予測モデル
3. 投資家に着目した予測モデル

市場の状態に着目した予測モデルには，過去の株価情報や，政治やマクロ統計に関わるニュース情報を用いた予測モデルがあり，テクニカルアプローチ [10–12]

とも呼ばれる。金融資産の性質に着目した予測モデルには、企業の財務情報やビジネスモデル・事業内容を用いた予測モデルがあり、ファンダメンタルズアプローチ [13, 14] とも呼ばれる。投資家に着目したアプローチは、市場参加者の心理や投資行動に着目した予測モデルで、金融経済学では行動経済学的アプローチ [15] と整理される。

金融市場の予測は、金融経済学の研究分野の一つとして多くの関心を集めているが、実務的にはただ予測を精緻に行うだけでは十分ではない。取引戦略を高度化するためには、ポートフォリオの最適化や、取引コストの最適化や、解釈・説明性の向上を考慮する必要がある。金融市場の予測は、取引戦略の高度化を目的にしたポートフォリオ最適化のための一つの手段である。より優れたポートフォリオを構築するためには、

1. 期待収益率を精緻に予測する
2. 分散・共分散を精緻に予測する
3. ポートフォリオ最適化手法を改善する

という3つのアプローチがある。金融市場の予測は、期待収益率を精緻に予測することと、分散・共分散を精緻に予測することを目的とした一つのタスクである。金融市場の予測を精緻に行うことで、ポートフォリオを構築する期待収益率や分散を精緻に予測することが可能になり、結果として取引戦略が高度化する。

金融市場の予測モデルの研究は、1. 着目し分析対象とする情報と、2. 研究の目的の2つの軸で図2.1のように位置づけ整理することができる。

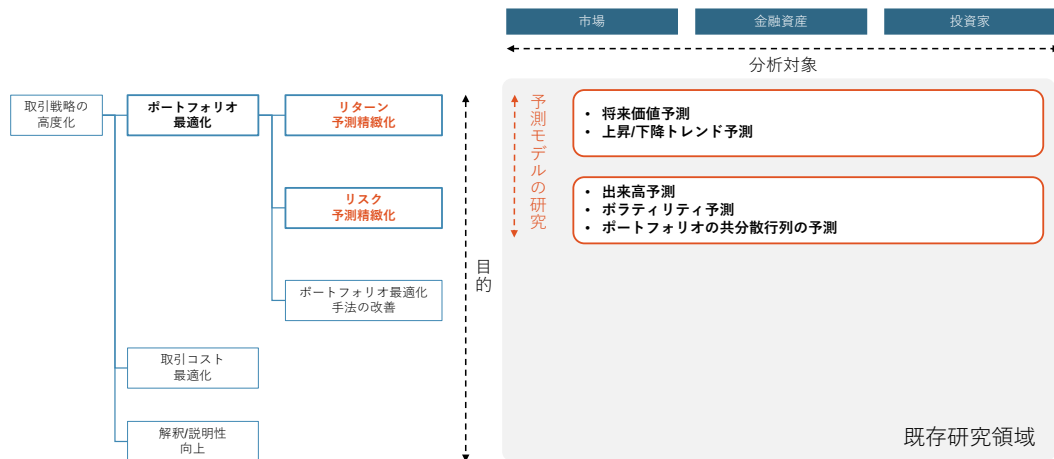


図 2.1: 金融市場の予測モデルの研究の位置づけ

2.1.2 効率的市場仮説と予測モデル

金融市場における価格形成は多様な因子が相互に影響し合う複雑な現象であるため、単純化するために幾つかの前提仮説を導入し、特定の切り口で理論化や予測モデルの構築を行う事が一般的である。前提仮説の違いや切り口の違いによって、理論や予測モデルの違いが生まれる。

次に、効率的市場仮説 [1] について述べる。効率的市場仮説とは、「全ての情報がすでに株価に織り込まれている」という仮説である。効率的市場仮説には程度で分類された以下の3つの説がある。

ウィーク型効率性

市場において現在価格は過去の情報を全て反映したものである

セミストロング型効率性

ウィーク型の仮説に加えて、新たに公開される情報が瞬時に価格に反映される

ストロング型効率性

インサイダーなどの一般的でない情報でさえ瞬時に価格に反映される

効率的市場仮説は、投資家が常に情報を瞬間的に「認知」し、正しい「判断」を行うという強力な前提が必要となる。

合理的期待形成仮説は、合理的であろうとする市場参加者は全体として、現時点での価格との比較で裁定取引を行いリターンを得ようとするため、近似的に市場は常に完全に情報的に効率的であるとする仮説である。個人単位ではノイズトレーダー [5] のような非合理的な行動をとる市場参加者が存在していたとしても、全体としては近似的に極めて合理的であるとすることによって、効率的市場仮説をサポートする。

効率的市場仮説や合理的期待形成仮説に基づき、ランダムウォーク理論 [6] が生まれた。ランダムウォーク理論は、株価の動きは過去の株価と独立で不規則的であり、予測をすることができないとする理論である。ランダムウォーク理論に基づくと、株価の予測そのものの可能性を否定する理論であるが、一方で、確率的には一定範囲内で価格形成がなされるのであれば、ある瞬間の静的な均衡状態をモデル化して分析可能になると考えられる。ランダムウォークする株価は正規分布に従うため、株式投資のリスクはリターンの分散として捉える事ができる。

リスクはリターンの分散という前提に基づくと、ポートフォリオのリスクも共分散を用いてモデル化する事ができる。CAPM理論 [7] は、市場ポートフォリオを導入し β を用いて単純化すれば、任意の金融資産の期待リターンを計算できるとする理論で、CAPM理論は基本的な資産価格モデルの一つであり、実用でも広く浸透している。[8] はCAPM理論のモデルとしての当てはまりの悪さを指摘し、より当てはまりの良いファーマ-フレンチの3ファクターモデルを提案した。提案者の1人である Fama はノーベル賞を受賞している。

効率的市場仮説や、効率的市場仮説を前提とするCAPM理論やファーマ-フレンチの3ファクターモデルは実用でも広く浸透している理論ではあるが、もし効率的市場仮説が常に正しいとすると、事実上市場の平均以上のリターンを求める事

が不可能であり、現在の価格が常に適正な価格であるから、過去の情報から将来の金融市場を予測する意味はなくなる。

しかし、実際には、株価が割高になるバブルの発生や暴落の発生の存在や、投資家は必ずしも合理的ではないという効率的市場仮説に対する反論も多く存在する。

値上がりした金融資産の価格がさらに上昇したり、値下がりした金融資産の価格がさらに下落すると言った市場が一方向に進みやすいというモメンタム効果や、時価総額が小さい小型株は大きい大型株よりも期待収益率が相対的に高くなる傾向があるとする小型株効果、簿価時価総額が高い金融資産ほど期待収益率が高くなる傾向にあるとするバリュー株効果などと言ったアノマリー [2] と言われる例外も指摘されている。RJ Shiller [16] は、過去の株式市場のデータから株価は常に合理的な価格ではないことを検証することで、効率的市場仮説を否定している。

また、人は合理的であろうとするが、人の認知・推論能力には限界があり、そうした限定合理性 [3] に基づき市場において意思決定を行うために、合理的市場形成仮説における市場全体の合理性が成り立たないともされている。実際には市場参加者の売買によって市場価格が決定しているが、その投資行動は個人の異なる効用に基づき決定しており、また市場参加者は常に入れ替わっている。そこで、市場参加者がどのような投資行動をとるかということに着目するような、行動経済学的なアプローチ [15] にも焦点が当てられ、議論されるようになった。

アノマリーの存在や限定合理性に基づく金融市場を予測と、市場平均を上回る期待収益 α の追求は重要な課題である。[9] では、予測とは何かという問いに対して、「予測かどうかは、主体がもつモデル及びその社会的な合意に依存する」と述べている。

金融市場の予測モデルとは、多様な因子が相互に影響し合う複雑な金融市場における価格形成において、人の認知限界を理由に単純化されていた前提仮説と現

実とのギャップや、アノマリーの存在に着目し、理論化する事で将来の価格形成を予測する手法であると考えられる。

2.2 過去の株価やニュース情報を用いて市場の状態に着目した予測モデルの研究

過去の株価やニュース情報を用いて市場の状態に着目した予測モデルの研究は数多く、機械学習を用いた手法 [17-27] も数多く提案されている。

これらの研究は、図 2.2 のように、

1. 時系列予測モデルの関数を改善する研究
2. モデルの入力として用いる情報ソースを拡張する研究
3. 取り扱う金融商品・予測対象を拡張する研究

の3つに分類して整理する事ができる。

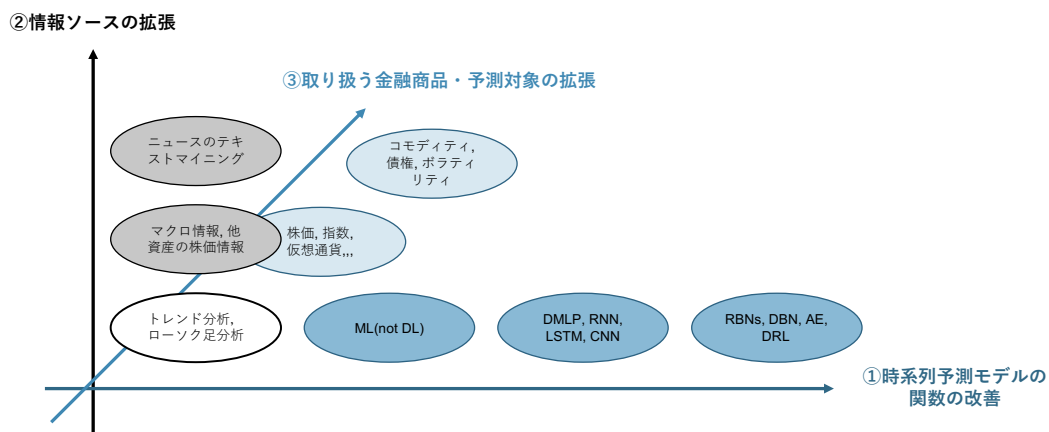


図 2.2: 過去の株価やニュース情報を用いて市場の状態に着目した予測モデルの研究の方向性の整理

2.2.1 市場の状態に着目した時系列予測モデルの研究

市場の状態に着目した時系列予測モデルの研究は、自己回帰和分移動平均モデル(以下, ARIMA モデル) [28] やベクトル自己回帰モデル(以下, VAR モデル) などの時系列モデリングをする手法 [29, 30] がある. 近年の機械学習分野の研究の発展に伴い, Time Delay Neural Networks を用いて株価のトレンド予測をおこなった研究 [23], Recurrent Neural Networks を用いて翌日の株価指数を予測する研究 [24] などが2000年以前に行われている. 2000年以後, Neural Networks と他の予測モデルを組み合わせるハイブリッド型のモデルが提案され始め, Neural Networks と遺伝子アルゴリズムと組み合わせた研究 [21] などがその一つである. また, サポートベクターマシンを利用した株価予測 [22] が精度をあげたことで話題を呼んだ. 近年の深層学習の研究の発展によって, Recurrent Neural Network の拡張として提唱された Long Short-Term Memory(LSTM) [18] を用いて株価指数予測をする研究 [19] をはじめ, 深層学習を用いた高度な関数を期待収益率を予測する時系列モデリングに組み込む [25–27] 手法も提案されている.

現実のポートフォリオ最適化のタスクに合わせて, 市場の状態に着目し期待収益率の予測をより精緻に行う時系列予測モデルの研究も続けられている. Deep Factor Model [31] は, 投資の意思決定に際して説明性や解釈性が求められる現実のポートフォリオの構築のタスクに合わせて, 予測に寄与するファクターを判断し説明することができる. 期待収益率の予測誤差の影響を少なくするように, ファクターモデルの共通因子の影響を削減しつつ, 正のリターンを得ることができる [32] という研究に基づき, ファクターモデルの残差項に着目した手法 [33] が提案されている.

2.2.2 モデルの入力として用いる情報ソースを拡張する研究

モデルの入力として用いる情報ソースを拡張する研究は, ファーマ-フレンチの3ファクターモデル [34] に代表されるクロスセクション情報や他の金融市場の情報を用いてモデリングするファクターモデルがある. 機械学習分野の研究の発展

に伴い、今までは直接利用する事が難しかったニュースの情報に対してテキストマイニングを行うことで、株価情報以外の情報もモデルの入力として利用できるようになった。例えば、イベントの情報 [35] や、報道されるニュース [36, 37] の情報に対して自然言語処理を行うことで期待収益率を予測する手法の研究が行われている。

2.2.3 取り扱う金融商品・予測対象を拡張する研究

過去の株価やニュース情報を用いて市場の状態に着目した予測モデルの研究の多くは、予測対象がS&P500(Standard and Poor's 500), TOPIX(Tokyo Stock Exchange Prices Indexes) や日経平均株価などの指数である。企業一つ一つの株価を予測しているような研究は、遺伝子アルゴリズムにより特徴量選択を行い、サポートベクターマシンを利用して株価を予測するハイブリッド型のモデル [17] や、新聞記事の情報を用いた LSTM モデルと株価情報を用いた LSTM を併用する研究 [20] がある。企業の固有な情報や特性なども株価に反映されることから、平均化された指数予測の研究とは性質が異なる。機械学習の発展によって、金融資産ごとに予測モデルを構築することも可能になってきている。

ポートフォリオの分散・共分散をより精緻に予測する手法の研究も進んでいる。過去の株価情報から標本共分散行列を計算することは、i) 株価の非線形なダイナミクスを捉えることができないこと、ii) ポートフォリオの構成する資産の数と共分散行列で推定する必要があるパラメタ数に対して、時系列のサンプル数が不十分になりがちで、低ボラティリティーの固有要因のリスクを過小に評価し、高ボラティリティーの固有要因のリスクを過大に評価する傾向があること、iii) 時系列普遍性を前提にしているが、実際はその限りではないこと の3つの課題が挙げられる [38, 39]。[39] では、標本共分散行列と β を単一指標として用いた共分散行列を加重平均することで、推定するパラメタ数に対してサンプリング数が不足するという課題に対処する手法を提案した。[40] では、ファクターモデルの残差項目に着目し、スパースな分散・共分散行列を導入することで、より精緻なポートフォリオの

分散・共分散を表現している。過去の株価データに対して、Graph Convolutional Neural Networks(GCN) [41] を用いることで企業間の関係を予測する手法 [42] も提案されている。他にも、平均分散アプローチ [43] の枠組みを超えて、収益率の期待値を用いずに、分散の予測結果のみを用いてポートフォリオを構築するリスクベースのポートフォリオの研究もなされている [44-48]。

2.3 金融資産が持つ特徴や性質に着目した予測モデルの研究

2.3.1 企業価値の推定と予測

金融資産が持つ特徴や性質に着目し企業価値を予測するタスクは、株価を予測するタスクとは性質が異なる。株価は、

$$\text{株価} = \text{株主価値} / \text{発行株式数}$$

で定義される。株価は企業間で比較する事はできないが、金融資産の価値を表す株主価値は比較可能である。そのため、金融資産が持つ特徴や性質に着目した予測モデルは、予測対象の性質と予測の目的を整理した上で用いる必要がある。

ファイナンスの分野において、「企業価値」は

$$\text{企業価値} = \text{負債} + \text{株主価値}$$

と定義される。

企業価値とは、上場企業だけに当てはまる概念ではなく、上場していない中小企業についても有効な概念である。一口に企業価値といっても、

1. 金融機関から見た企業価値
2. 組織再編性及び相続時などにおける事業関係者にとっての企業価値
3. 企業買収、売却の関係者にとっての企業価値
4. 投資家にとっての金融商品としての企業価値

といった様々な視点から捉えることができる。(1) 金融機関から見た企業価値については、多くの企業は金融機関から資金調達を行なっているが、金融機関はその企業が投資先としてどのようなリスクがあり、どれほどの価値があるのかを見定める必要がある。(2) 組織再編性及び相続時などにおける事業関係者にとっての企業価値については、企業関係者は自社の価値を知ることなしに組織再編性や相続を行うことは不可能であるし、また自社の価値を正しく推定することで競合他社との比較や自社の経営改善などを行うことも可能になる。(3) 企業買収、売却の関係者にとっての企業価値については、企業を買収もしくは売却するかの判断には、買収もしくは売却による定性的なメリット・デメリットだけでなく、企業価値を定量的に推定することが欠かせない。また、(4) 投資家にとっての金融商品としての企業価値については、投資家にとって、金融商品の価値を推定することは極めて重要なことであり、特に将来どのような価値として評価されるか、つまり価値が上がるか下がるかは利益に関わる本質的な問題である。

企業価値を推定する目的もその性質によって異なる。(1) 金融機関から見た企業価値及び(2) 組織再編性及び相続時などにおける事業関係者にとっての企業価値においては、主として「企業価値」を知ることが最終目的であるのに対し、(3) 企業買収、売却の関係者にとっての企業価値及び(4) 投資家にとっての金融商品としての企業価値において企業価値を推定することの最終目的は「将来の株主価値」または「将来株価」を推定することである。

そして、推定された企業価値は、推定対象企業が上場企業か非上場企業かによってその性質を異にする。対象企業か非上場企業であった場合、推定価値を検証する手段がなく、他に参考になる値が存在していないので、推定値そのものに価値がある。一方で対象企業が上場企業であった場合、株価という形で常に市場の評価値が出ているために、市場の評価との比較が常に重要であり、推定価格が「将来の市場の評価値」であることが求められる。

本稿では、金融資産が持つ特徴や性質に着目した予測モデルの対象を(3) 企業買

取、売却の関係者にとっての企業価値及び(4)投資家にとっての金融商品としての企業価値とする。

2.3.2 経済学的ファンダメンタルズアプローチ

既存の企業価値推定モデルとして用いられている経済学ファンダメンタルズアプローチについて整理し、それらの抱えている問題を考察する。経済学的ファンダメンタルズアプローチには、アセットアプローチ、マーケットアプローチ、インカムアプローチと主に3つの手法が存在する。

アセットアプローチ

アセットアプローチは、対象企業の貸借対照表に記載されている資産と負債の差額として計算される純資産をもとに企業価値を推定する方法である。アセットアプローチはさらに下記のような手法に分類される。

簿価純資産法

貸借対照表に計上されている資産・負債に粉飾・誤謬（例えば、減価償却の不足分など）等の修正を加え、資産から負債を控除した純資産額をもって株式評価するもの

修正簿価純資産法

簿価純資産に一部の資産・負債の含み損益を取り込んで評価する方法。有価証券や土地などで大きな含み損益が認められるものについて時価の取り込みを行う。

時価純資産法

資産・負債の各項目について時価評価し、時価純資産を算出し株式評価する。売上債権の滞留債権や棚卸資産の滞留在庫を評価減したり、退職給付債務や損害賠償等の簿外処理されている可能性のある負債をオンバランスするなどして、時価純資産を算定する。時価純資産法は、個別項目ごとに時価評価し、それらを積み上げて時価純資産を求める。

アセットアプローチは、単純化されたアプローチで客観的であるという利点があるが、企業価値は将来キャッシュフローによって決まるという古典的なファイナンス分野の概念と合致していないことや、将来の予測になっていないことなどが問題点としてあげられる。

マーケットアプローチ

マーケットアプローチは、市場価格を基礎にして対象評価会社の株式の評価を行う方法である。第三者間取引が最も公正な価格であるという前提に立ち、市場において実際に取引されている価額を用いるものである。マーケットアプローチはさらに下記のような手法に分類される。

類似公開企業比準(比較)法

上場(公開)している類似企業の各種指標の株価に対する倍率を参照して、評価対象会社の株式評価を行う方法。評価対象企業が非上場企業であるときにDCF法と合わせて頻繁に用いる方法で、マーケット・アプローチの代表的な評価方法である。

市場株価法

評価対象企業が上場企業である場合に用いる方法。採用する株価は、前日終値、終値1カ月平均値、終値3カ月平均値、終値6カ月平均値などがあり、評価目的に合わせて単一もしくは併用して評価額を求める。

類似取引法

類似した取引を参照し、参照した取引における倍率等を用いて評価する方法。欧米等ではM&A取引などで用いられているが、日本では詳細に公表されている取引実績がないため類似取引のデータベース化がすすんでおらず、評価実務ではあまり一般的に採用されていない。

過去取引事例法

評価対象企業の株式が過去に頻繁に売買されている場合、過去に取引された際の価格を参照する方法。

マーケットアプローチの最大の利点は実際の取引価格を参照している点である。しかし、市場の価格形成の特性を十分に理解し、評価目的に照らして選択しなければ、誤った評価額を導く可能性がある。

インカムアプローチ

インカム・アプローチは、将来期待される一連の経済的利益ないしキャッシュ・フローを、その実現のために想定されるリスク等を反映した割引率で現在価値に割り引いて企業価値を算定する方法である。インカム・アプローチは、M&A等における評価額をはじめ、様々な場面で利用される最も基本的な評価方法であり、最も代表的な評価方法であるDCF(ディスカунテッドキャッシュフロー)法は、将来にわたって得られるキャッシュフローを現在価値に割引いた額となるため、ファイナンス理論に最も忠実な方法とも考えられている。インカムアプローチはさらに下記のような手法に分類される。

DCF(ディスカунテッドキャッシュフロー)法

将来キャッシュフローを、リスク等を勘案した割引率によって現在価値に割り引いた金額を評価額とする評価方法。ファイナンス理論に最も忠実な評価方法といえ、実務的に様々な場面で利用されている。ただし、将来キャッシュフローや割引率の設定に主観的な判断が介在する可能性も高く、評価額の妥当性を十分に吟味する必要がある。

収益還元法

会計上の予想利益を資本還元率によって割り引くことで企業価値を算定する方法。収益還元法は一定の成長率をもって評価する方法であるためDCF法に比べて評価額が硬直的となりやすく、予想利益が変動することが想定される場合には効果的な評価方法とは言えず、株式評価報告書ではあまり採用されない。ただし、一定の成長率であることがある場合には合理的な評価方法であり、またコストをあまりかけずに評価する場合にはDCF法の簡便的な評価方法として採用することができる。

配当還元法

配当金を株主資本コストで割り引くことで株主価値を直接評価する方法。DCF法や収益還元法が企業価値を評価し、そこから株主価値（すなわち株価）を評価する方法であるのに対し、配当還元法は株主価値を直接求めることができる。ただし、配当金が見込めない企業（多額の繰越欠損金を抱えている、ベンチャー企業である）や日本企業のように低配当政策を採用する企業の場合、低い評価額となる可能性がある。

2.3.3 経済学的ファンダメンタルズアプローチの課題と研究の位置づけ

既存の経済学ファンダメンタルズアプローチの特徴として、インカムアプローチのDCF法などは上場企業にも非上場企業にも適用できるということがある。それはファイナンス理論に従って演繹的なアプローチによって導出しているからである。

しかし、いずれのアプローチも、過去、現在、未来という時間軸に従った分析になっておらず、現在の状態を基準にして現在値の推定及び将来予測を行なっているにすぎない。また、マーケットアプローチのように他企業との比較を行う手法もあるが、比較企業の選択が同業界内であったりと範囲が限定されかつ恣意的であることが多い。経済学的ファンダメンタルズアプローチは、長い時間軸、及び企業間の関係を考慮するアプローチではない。

さらに、既存の企業価値推定モデルとして用いられている数式モデルが、時代の変化に対応できていない可能性がある。例えば、DCF法の割引率として用いられることの多いWeighted Average Cost of Capital(以下、WACCと呼ぶ)の株式資本コストは、株のリスクを市場の動きから予測しようとするCAPM理論 [7] を用いて導入されることが多い。しかし、CAPM理論が提案されたのは1960年代であるが、当時から日本の10年国債の利回りは大きく減少し、日本の株式市場の状態は大きく変化している。日本の株式市場の現状が過去に提案された数式モデルに適していない可能性がある。

金融資産が持つ特徴や性質に着目した予測モデルの研究は、図2.3のように、

1. 既存のアプローチにおける手法の一部に統計・機械学習手法の導入する研究
2. 財務情報やビジネスモデル・事業内容などの情報ソースを拡張する研究

の方向性に整理できる。データ構造の画一性が低い情報に対して、ヒトが複数資産に着目した帰納的推論を行うことは難しく、機械学習を用いた手法の有用性が期待される。

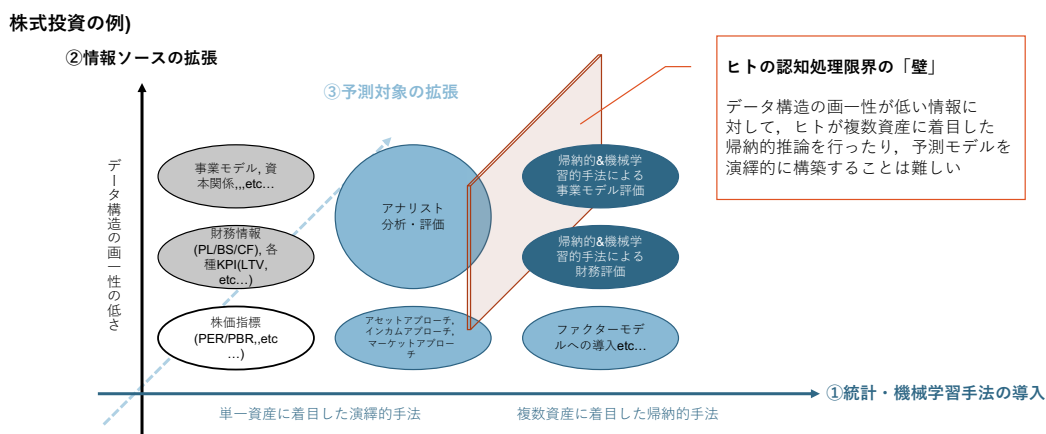


図 2.3: 金融資産が持つ特徴や性質に着目した予測モデルの研究の方向性の整理

2.4 ソーシャルメディアを用いて投資家に着目する予測モデルの研究

2.4.1 行動経済学とソーシャルメディアによる予測モデルの可能性の拡張

行動経済学では、個人の心理状態が個人の行動や意思決定に強い影響を及ぼすとされている [49–51]。さらに [52] では、金融市場に関わる意思決定は、感情や心理が強く影響していることが示されている。

ソーシャルメディアが普及する以前は、金融市場において市場参加者や世間がどのような感情を抱き、何に注目しているかを獲得することは限定的であった。ソーシャルメディアの登場によって、ユーザーの興味関心や心理状態を連続的に、リアルタイムで獲得することができるようになったことで、ソーシャルメディアを

用いた金融市場の予測可能性を広げた [53]. 実際, ソーシャルメディアを用いて市場参加者についての興味関心や心理状態を獲得し, そうした情報を用いた金融市場を予測・説明しようと試みる研究がなされている [53-55].

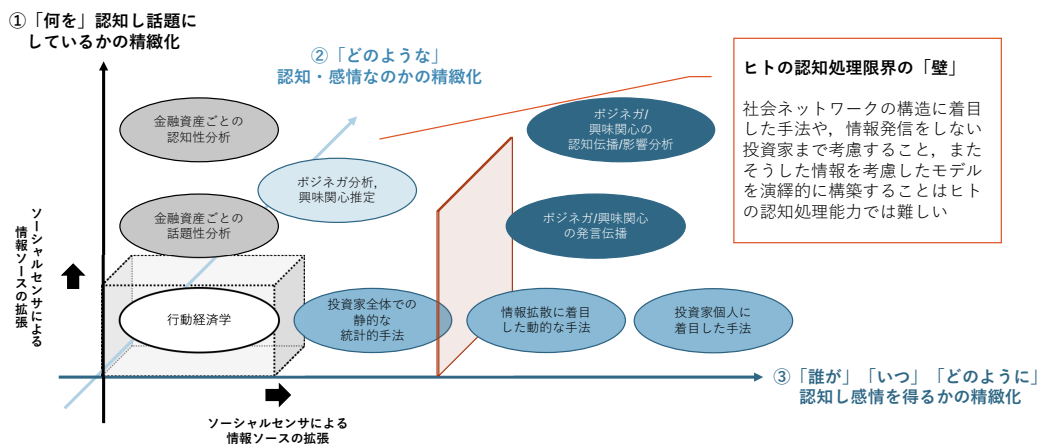


図 2.4: ソーシャルメディアを用いて投資家に着目する予測モデルの研究の方向性の整理

ソーシャルメディアを用いて投資家に着目する予測モデルの研究の方向性は, 図 2.4 のように,

1. 投資家が「何を」認知し話題にしているかの精緻化する研究
2. 投資家が「どのような」認知・感情なのかの精緻化する研究
3. 投資家の「誰が」「いつ」「どのように」認知し感情を得るかの精緻化する研究

と整理できる.

ソーシャルメディアによって情報ソースが拡張され, 金融市場の予測可能性は広がったが, 社会ネットワークの構造に着目した手法や, 情報発信をしない投資家まで考慮すること, またそうした情報を考慮したモデルを演繹的に構築することはヒトの認知処理能力では難しく, 機械学習を用いた手法の有用性が期待される.

2.4.2 仮想通貨の登場と投資家心理を考慮することの重要性

近年、ブロックチェーンという技術に基づいた仮想通貨の存在が話題になっている。Kristoufek [56] は、仮想通貨を以下のように定義している。

”A digital currency can be defined as an alternative currency which is exclusively electronic and thus has no physical form. It is also not issued by any specific central bank or government of a specific country and it is thus practically detached from the real economy.”

仮想通貨には、ビットコイン、イーサ(イーサリアムプラットフォームの内部通貨、以後一般的な呼び方に習いイーサリアムと呼ぶ)、リップルなど様々な通貨があり、ビットコインは2017年初頭では\$900ほどであったのに対して、2017年12月には\$20,000 近くまで価格が上昇した。ここでは、特に2018年11月時点で仮想通貨の中でも時価総額が大きいビットコインについて整理することで、金融商品としての性質を理解する。

ビットコイン [57] は、ブロックチェーン¹という Peer to Peer ネットワークを用いた分散処理による時系列データの記録技術を用いた仮想通貨の一種であり、2021年現在仮想通貨で最も時価総額が高い。

国家の信用に基づいた価値定義がされている法定通貨と異なり、ブロックチェーンという非中央集権的な技術によって価値が定義されている点が仮想通貨の特徴である。ブロックチェーンは、1つ前のブロックのハッシュ値の Prev Hash, ハッシュ値を計算するための一時的な暗号の Nonce, トランザクションなどの情報をハッシュ化し固定長にした Root Hash という要素で構成されるブロックを、不可逆性を持つハッシュ関数を再帰的に用いることによって連結し、データの不可逆的な記録構造を持つ仕組みである。最も連結されているチェーンが真のデータとし、ハッシュ関数の持つ非対称性²を利用して、新しいブロックを追加するためには多

¹ [57] においては、ブロックチェーンという言葉は使われておらず、Timestamp Server という表現がなされている

²ハッシュ関数は、その計算は簡単であるが、逆関数を求めることは難しい

大なコストを必要とする Proof-of-Work [57] という仕組みを採用することでデータの改竄を防いでいる。

ビットコインは、ブロックチェーンと Proof-of-Work 技術を用いることで、非中央集権的であるにも関わらず改竄が非常に困難なデータであるものの、Proof-of-Work の新しいブロックを追加するために多大なコストを要するその性質から、決済には約 10 分の時間を要し、法定通貨などに比べると決済に向いていない。そうしたことから、ビットコインの対改竄性などの以下の性質から、「デジタルゴールド」などと呼ばれ、金と同様の性質があると考えられている。

- ビットコインには総発行数があらかじめ決められており、希少価値がある。
- 偽造ができない。
- 獲得にはコストがかかる。
- 劣化しない。

金融商品としての金は、以上のような性質から安全資産であると認識され、その価値は他の有価証券などの価値と逆相関をなすことが知られている。

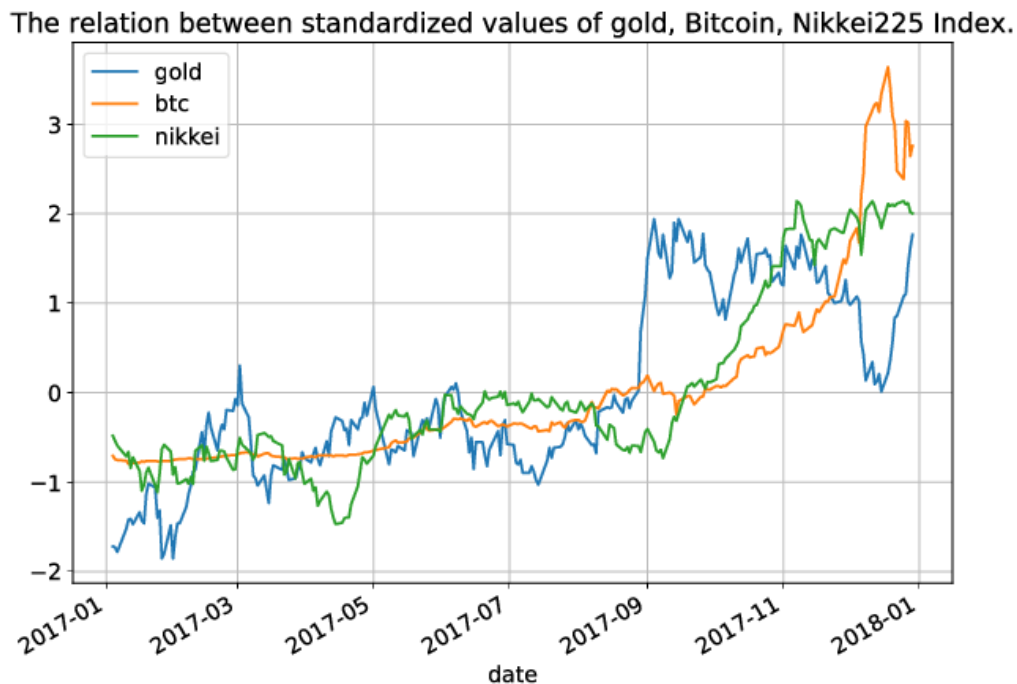


図 2.5: 2017 年における金, ビットコイン, 日経平均指数の価格推移の関係. 金と日経平均には逆相関が見られるが, ビットコインと日経平均にはそのような関係はない.

しかし, 図 2.5 のように, ビットコインと日経平均指数の価格変動には逆相関は見られず, 実際には金のような安全資産としての役割が期待されていないことがわかる.

ビットコインをはじめとする仮想通貨はマクロ経済に基づくファンダメンタルズ要素がないとされるため [56], 将来のキャッシュフローモデル [58], 購買力平価 [59, 60], および金利パリティ [61, 62] のような既存の経済理論や金融理論によって仮想通貨市場における変動を予測したり, 説明することは困難である. 仮想通貨市場の変動は, 投機的な目的における需要と供給によって決まっている可能性があり, 投機的な目的における需要と供給の関係は, 市場参加者の心理に大きく左右されると推察される.

仮想通貨のようなファンダメンタルズの価格決定要因を持たず、相対的に市場参加者の心理が重要であると考えられる金融商品の出現により、市場参加者の心理状態と金融市場における関係の研究の重要性は増している。

2.4.3 ソーシャルセンサの研究

2000年代後半、ツイッターやフェイスブックと行ったソーシャルメディアが登場した。スマートフォンの普及と共に、人々は、実世界に情報端末をもち出し、そこで起こった出来事や感じたことをソーシャルメディアに投稿するようになったため、今までデータとなっていなかったような社会や実世界で起きている事象でさえコンピューターに記録されるようになった。そうしてソーシャルメディアは、社会現象の実態やその発生メカニズムを解明しようとしきた社会学において、貴重なデータ資源となり、社会ネットワークの分析や実世界の観測や予測を行うためのソーシャルセンサ [63] として機能している。

ソーシャルメディアの登場によって、今まではデータ化することが難しかった実社会におけるネットワークが記録されるようになり、実社会における人間関係のネットワークの解析が可能になった。実際に、MSN メッセンジャーの友人関係を調査した研究 [64] や、フェイスブックの友人関係を調査した研究 [65] がある。

フェイスブックなどのソーシャルメディアは、双方向ネットワークであるのに対して、ツイッターはエッジ方向がある有向グラフであるという特徴を持つ。ツイッターの有向グラフを調査した研究 [66] では、フォロー関係にあるユーザーのうち、相互リンクである割合は22.1%にとどまり、大半が一方方向的な関係にあることが示された。このことから、ツイッターのユーザーネットワークは、実世界の人間関係を表した社会ネットワークというよりも、人々の興味の関係を表した情報共有ネットワークであるとみなされるようになった。写真を投稿するソーシャルメディアであるインスタグラムも、ツイッターと同様に有向グラフのネットワークをもち、インスタグラムのネットワークを調査した研究でも、ネットワークの関係性は人々の好みや関心に従っていることが示唆されている [67]。さらに、ツ

ツイッターのユーザーのネットワークにおける任意の二人の隔たりの字数の平均は、一方方向のネットワークであるにも関わらず、フェイスブックのネットワークの研究 [65] で示されていた6次よりも小さい4.12であり、情報拡散を行うメディアとしての強力さを示すこととなった。

特にツイッターやインスタグラムのような有向グラフを持ち、強力な情報拡散の機能を持つソーシャルメディアの登場によって、人々は実社会の人間関係に加えて、人々の興味関心に従った新たな社会ネットワークを形成した。

ソーシャルメディアにおいて新たな社会ネットワークが形成されたことによって、実世界の関係性や人々の興味関心に基づく社会ネットワークの分析がより精緻に行えるようになっただけでなく、ユーザーの投稿やそれに関する反応から、社会で起こりつつあるイベントやトレンド、現象などを検出することが可能になった。

ソーシャルメディアでは、今までの情報媒体に比べて、より即時的に、より広範囲の社会イベントが記録されるようになっている。これは、ユーザーが常時スマートフォンを持ち歩くようになったこと、そしてツイッターのような有向グラフの社会ネットワークをなすソーシャルメディアでは特に、ユーザーの結びつきに相手の認証を必要としないため、より気軽に情報を発信するようになったことが背景にある。ソーシャルメディアにおける情報は、実世界で何が起きているのか、そしてそれに対して人々がどう反応しているのかを示している情報であり、社会で起こりつつあるイベントやトレンド、現象などを検出するソーシャルセンサとして機能している [9]。ソーシャルセンサを通して得られた出力から、統計処理などによって実世界の観測や予測を行うことができ、ソーシャルセンサを用いた実世界の観測や予測に関する研究がいくつかなされてきた [53, 54, 56, 63, 68, 69]。

例えば、ソーシャルセンサによって、地震が起こった時に、ツイッターにおける地震に関する投稿から地震を検知することができる [63]。将来の映画の収益をツイッターから得られる情報から予測したり [69]、インスタグラムの投稿からウェブ上で人気になったり関心を集める事象を予測することができる [67]。選挙における投票では、ツイッターから得られる情報を用いることで当選者を予測する研

究 [68, 70] がなされている。経済指標をツイッターから得られるセンチメント情報によって予測するモデルも提案されており [53–55], Googleトレンドとビットコインのマーケットの相関の存在なども示されている [56].

近年はソーシャルメディアを用いた観測と予測を、現実のタスクにおいてより精緻に行うための研究が続けられている。[71]では、分析対象となるコンテンツの新規性の高さゆえ、観測頻度が低く、抽出が難しいコンテンツを抽出する手法を提案している。[72]では、ソーシャルメディアにおいて特定の領域の分析を行うには、サブグラフを特定する必要があるが、ソーシャルメディアのような動的で多次元的なネットワークにおいてそうしたサブグラフを効率的に抽出する手法が研究されている。また、ユーザの人気, social rank を考慮した network embedding の手法 [73] は、大規模なネットワーク構造を持つソーシャルメディアの活用可能性を高めた。

近年、「インスタ映え」という、インスタグラムに投稿するために写真にこだわる現象から生じた言葉が話題になり、他人から「いいね」をもらうようなインスタグラムに載せる画像を撮るためにレストランに行ったり、旅行に行ったりする消費行動が報告されている。また、金融市場においても、市場参加者は他の市場参加者のツイッターにおける投稿から取得した情報に基づいた投資判断を行い、ある特定のユーザーのツイートによって株価が変動している場合がある可能性が述べられている [74].

ソーシャルメディアが普及するにつれて、人々が双方向の友人ネットワークによって構成される実社会だけでなく、ソーシャルメディアに形成される新しい社会ネットワークをより重要視するようになり、ソーシャルメディアにおける新たな社会ネットワークは、我々の実世界の社会に対して働きかけているというソーシャルアクチュエーターとしての機能を持っているという仮説が成り立つ。

ソーシャルセンサに関する既存研究は、実世界の観測と予測の可能性を広げた

一方で、ソーシャルメディアを単なる実世界の射影や、実世界を観測・予測したりするための情報源としてしかみなしておらず、ソーシャルメディアに形成される社会ネットワーク内での相互作用が、実世界に働きかける可能性について考慮した研究は行われてこなかった。実世界におけるイベントがソーシャルメディアのネットワーク構造や内部のコミュニケーションに影響を与えることを分析した研究 [75] や、ソーシャルメディアのユーザの時系列状態を隠れマルコフモデル (以下 HMM と呼ぶ) を用いて表現し、ユーザのソーシャルメディア内における行動を予測する研究 [76] はなされている。しかし、ソーシャルメディアに形成される社会ネットワークにおいて、投稿の閲覧に伴う影響関係や、ユーザの認知・心理状態の変化が、実世界に対して働きかけていることを考慮し、実世界の観測と予測を行う研究はなされていない。

また、ソーシャルメディアにおけるユーザの影響関係をモデル化する際、直接観測することのできない影響関係をモデル化することは難しく、そうした研究は少ない。ソーシャルネットワークにおける影響関係をモデル化する既存手法として、Linear Threshold Model (以下 LTM と呼ぶ) [77] がある。しかし、LTM では、ユーザ u_i がユーザ u_j から受けた影響度を直接知ることができないため、影響度をネットワークグラフの指標などで代替して表現するが多かった [78]。金融市場に対して影響を及ぼすソーシャルメディアの相互作用をモデル化する場合、金融資産の性質やユーザの興味関心といった性質を反映し、直接観測することのできないユーザの認知や心理状態の変化をより精緻に考慮する必要がある。

2.4.4 information diffusion の研究

ソーシャルメディアの社会ネットワークにおける影響関係が、将来の実世界の社会における事象に対してどのように作用するかを研究する際、ソーシャルメディアの人々の興味関心に従った新たな社会ネットワークにおいて存在するユーザーの影響関係は、既存研究の情報拡散の仕組みや問題設定と類推できる。information diffusion の関連研究を整理し、ソーシャルメディアから実世界への作用を考慮し

た場合の実世界の観測と予測との問題設定の違いを明確にする。

information diffusion は, [78] で図 2.6 のように以下の4つのタイプに分けて説明されている。

- Herd behavior
- Information cascades
- diffusion of innovation
- epidemics

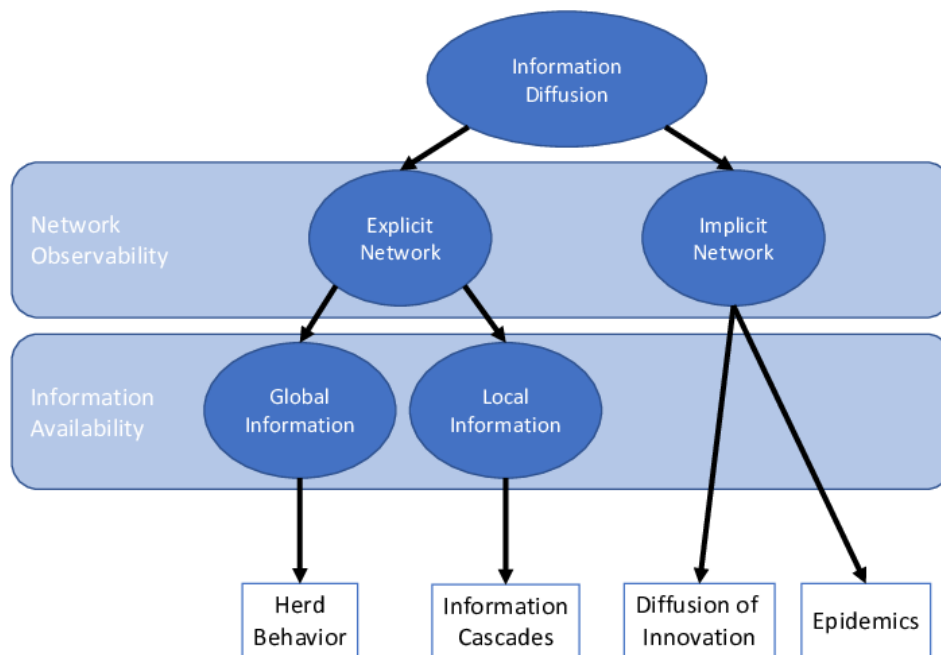


図 2.6: information diffusion の4分類

Herd behavior

Herd behavior とは、明示的な社会ネットワークにおいて、ある個人がネットワーク全体を観測することができ、社会ネットワーク全体から得られた情報に基づいて意思決定や行動を行う過程である [78]。一般的に、Herd behavior は以下の4つの要件を満たす必要がある [78]。

- 意思決定が行われる。
- 意思決定は連続的である。
- 意思決定は全く思考停止というわけではなく、個人の何かしらの情報にも基づくものである
- いかなる直接的な意思情報のやり取りは行われない。個人は他の個人情報を知ることはできないが、その行動から他人が何について知っているかは推論可能である。

Herd behavior は、ベイズモデルなどでモデル化され [79]、金融市場においても Herd behavior が起こり、ボラティリティの増幅や市場の不安定化、非合理的な投資行動がなされていることが研究されている [80–83]。

こうした研究は、マクロ的な視点において、社会ネットワークにおける information diffusion によって、金融市場が変動することを示唆している一方で、ソーシャルメディアのように隣接したノードに強く影響を受ける形での影響関係とは異なり、ソーシャルメディアから金融市場に対する作用の研究とは問題設定が異なる。

Information cascades

ソーシャルネットワークにおける information cascades は”An information cascade describes the process of diffusion when individuals merely observe their immediate neighbors” [78] と定義されている。Herd behavior との違いは、ユーザーは直接隣接するユーザーからの情報のみしか観測することができない点である。ソーシャルメディアでユーザーが観測する情報は、ソーシャルメディアの社会ネットワー

クにおいて直接友人関係にあるユーザーからの情報に限定されるため、ソーシャルメディアにおける information diffusion の研究は、一般的に information cascades の研究として扱われることが多い。

ソーシャルメディアにおける information cascades の研究として、情報の送信者と受信者の関係を獲得することを試みた Embedding-IC [84] や、大規模ネットワーク構造のベクトル化および説明をする LINE [85, 86] が提案されたり、将来の cascades の予測性について議論されている [87]。情報伝播やネットワーク構造の表現手法が、対象としているネットワークのドメイン知識に依存するような属人的な手法であることが問題としてあげられており、情報伝播やネットワーク構造を End-to-end のニューラルネットワークを利用することによってサンプリングし獲得する DeepCas [88] をはじめ、様々な手法が提案されている [84, 87-92]。

しかし、これらの研究は直接観測される情報拡散を観測・予測対象としており、投稿に対するいいねやリツイート、リプライといった直接観測可能な information flow のみしか取り扱っておらず、ソーシャルメディアにおいてアクションがないユーザに関しては考慮していない。ソーシャルメディアに形成される社会ネットワークから金融市場への作用を考慮し、金融市場の観測・予測を行う場合、直接観測することができないような、他者の投稿を閲覧することによるユーザ間の影響関係や、時間を通して変化・蓄積されるユーザの認知・心理状態をモデル化することが必要となる。

Diffusion of Innovations

Diffusion of Innovations [93] は、Innovation がソーシャルネットワークにおいて拡散する現象のことであり、ソーシャルメディアにおいて新しい音楽が大流行したり、一部のニュースが何回もリツイートされることが例として挙げられる [78]。Herd behavior や Information cascades の研究と異なる点は、ソーシャルネットワークにおける相互作用を観測することができず、情報拡散を俯瞰的に捉えているこ

とである。

[93]において、Diffusion of Innovations は以下の5つの過程に整理されている。

1. Awareness: 個人が特定の innovation について認知し始めているが、その innovation に対する情報は不完全である。
2. Interest: 個人がその Innovation に対して興味関心を抱き、もっと情報を知りたいと思う。
3. Evaluation: 個人が、自分がその innovation を使うことを想定し、実際に使うかどうか判断する
4. Trial: 個人が、試しにその innovation を使ってみる。
5. Adoption: 個人がその innovation を使い続けるかどうか決定する。

Diffusion of innovations では、情報拡散を俯瞰的に捉えており、微分方程式によってモデル化されている。

$$\frac{dA(t)}{dt} = i(t) [P - A(t)] \quad (2.1)$$

$A(t)$ は、時刻 t までに innovation に対して、上述した Adoption の状態にある人数を表し、 $i(t)$ は時刻 t の拡散係数、 P は潜在的に Adoption の状態になりうる人数を表す。式 2.1 は、時系列における Adoption の状態にある人数の変化率が、innovation の特性を意味する拡散係数 $i(t)$ に依存することを意味している。 $i(t)$ を表す手法はいくつか存在するが [94]、線形の数式モデルとして、以下のように表すことができる。

$$i(t) = \alpha + \sum_{i=t_0}^t \alpha_i A(i) \quad (2.2)$$

α はソーシャルネットワーク外部からの影響を表す変数、 $\alpha_i A(i)$ はソーシャルネットワーク内における影響を表す変数と考えることができる。

確かに、Diffusion of innovations に関する研究は、式 2.2 のように、ソーシャルネットワーク内における影響関係が innovation の拡散に寄与することを考慮している。しかし、ソーシャルネットワークにおける相互作用を観測することができないという前提に立ち、情報拡散を俯瞰的に捉えたものに過ぎないし、拡散係数 $i(t)$ は innovation の質に依存するものだという考え方に基づいており、ソーシャルメディアにおいて形成されている興味関心に基づく新しい社会ネットワーク内での相互作用に着目していない。

Epidemics

Information diffusion における Epidemics は、Information cascades や Herd behavior と異なり、Diffusion of innovation と同様個人間の関係が明示的でない場合の情報拡散の現象である。Epidemics がそれらの情報拡散の現象と大きく異なる点は、社会ネットワークにおける個人が情報拡散を意図してコントロールできないという点である [78]。

Epidemics において、社会ネットワークにおける個人は以下の3つの状態のどれかに該当するとされる。

1. Susceptible: 個人が特定の情報や病気に対して感染する可能性がある状態。簡単のため、感染は感染者からのみしかおこらないと仮定する場合が多い。
2. Infected: 感染している状態。
3. Recovered: 感染から回復し、完璧なもしくはある程度の感染に対する抗体を持っている状態。

[95] の研究では、上記の Susceptible, Infected, Recovered の状態を微分方程式を用いてモデル化し、SIR モデルなどと呼ばれている。近年では、Hawkes 過程 [96] と SIR モデルを組み合わせたモデルが提案されている [97]。

本研究では、ソーシャルメディアの形成する社会ネットワークにおける相互作用に着目し、それは個人の意図に基づくものであるため、Epidemic の情報拡散と

は問題設定が異なる。

2.4.5 投資家の影響関係を表す研究

社会ネットワークを介して、アイデアや意見、デマに至るまで非常に多様な情報が急速、かつ大規模に拡散し、われわれの日常生活に多大な影響を与えつつある [98]。Information cascades の研究では、リツイートやリプライといった観測可能な information flow による情報拡散を取り扱っているのに対して、本研究では投稿の閲覧といった直接観測することができない影響関係を取り扱う。そのような情報の受信者に着目し、影響関係をモデル化する既存手法として、Linear Threshold Model [77](以下 LTM と呼ぶ)がある。

LTM では、ユーザ u_j からユーザ u_i に対する影響度 $s_{j,i}$ を用いて、ユーザ u_i が受ける影響 r_i を以下の式のように表される。

$$r_i = \sum_{u_j \in N_{in}(u_i)} s_{j,i} \leq 1 \quad (2.3)$$

ただし、 $N_{in}(u_i)$ はユーザ u_i の隣接するユーザーを表す。LTM では、全てのユーザ u_i に対して、区間 $[0, 1]$ から一様ランダムに閾値 θ_i を選択し、割り当てる。そして、時刻 t で非アクティブであるユーザ u_i は、影響度の総和 r_i が θ_i を上回った時、時刻 $t + 1$ でアクティブになる [77]。

$$r_i \geq \theta_i \quad (2.4)$$

しかし、LTM では、影響度 $s_{j,i}$ を直接知ることができないため、影響度 $s_{j,i}$ をネットワーク指標などで代替して表現することが多く [78]、相互に複雑に関連し、かつ直接観測されないソーシャルメディアの投稿の閲覧に伴う影響関係と、時系列において変化・蓄積されるユーザーの認知や心理状態をモデル化することができていなかった。

第3章 深層学習を用いたアンサンブルモデル による株主価値推定モデルの提案

本章では、株式市場における資産の期待収益率の予測を目的に、市場に着目し過去の株価情報を利用した予測モデルにとどまらず、投資対象である金融資産それ自身に着目し財務情報を活用した予測モデルを提案する。

3.1 はじめに

株価は多様な因子が相互に影響しあう複雑な現象の結果であり、予測の可能性やモデリングの手法に関して、多様な分野で学問的関心を集めてきた。株式分析の手法は、ファンダメンタルズ分析 [13, 14] とテクニカル分析 [10–12] に大分され、[99] では、「テクニカル分析は市場の動きの研究に専念するのに対し、ファンダメンタルズ分析は価格を上下させたり、あるいは同じレベルにとどめさせたりする原因となる、需要と供給関係に焦点をあてる」としている。

株式市場分析の研究にも機械学習の手法が用いられている。大規模なデータからパターンを学習することができる機械学習は、テクニカル分析に基づいた短期的なトレーディングのためのトレンド予測に用いられることが多い。

しかし、健全な市場を形成し、より多くの投資を促進するという観点からすれば、中長期的な投資が重要な役割を担っている。1600兆円を超える日本の個人資産をより多く市場に流入させるためには、公正でかつ中長期的な投資を推進することが必要である。実際に、[100]によれば、金融庁は「早耳情報に基づく短期的なトレーディングではなく、中長期的な視点に立って投資を行うという投資家の意識変革を促す」という目的で、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。

中長期的な投資を行うためには、企業の財務諸表、経営状態、競争優位性などから、市場における価値に着目するファンダメンタルズ分析も重要である。なぜなら、上場企業の株価は、中長期的には市場原理に基づき、市場における企業の価値の評価によって決定しているからである。例えば、類似する A 社と B 社があるとして、A 社がより良い業績を発表した場合、B 社の投資価値そのものが下がったわけではないのに、A 社の価値が B 社よりも高いとして、B 社の株が売られ A 社の株が買われることが考えられる。また、C 社の業績自体は向上している場合でも、市場が期待していた水準ではないと認知された場合、実際の企業の価値が、期待されていた企業の価値よりも低いために C 社の株価が下がるということもある。よって、中長期的な投資においては、例え最終的な推定対象が株価であっても、株価そのもののトレンドの動きを直接推定するテクニカル分析だけでなく、市場における評価価値である株主価値に着目したファンダメンタルズ分析が重要である。

しかし、トレンドを推定するようなテクニカル分析に加えて、株主価値を推定するファンダメンタルズ分析を組み合わせた機械学習による手法は筆者の知る限り存在していない。確かに、テクニカル指標だけでなくファンダメンタルズ指標を扱う機械学習モデルは提案されているが、それはファンダメンタルズ指標を予測対象企業に帰属する入力変数として用いているにすぎず、需要と供給の関係を考慮し、株式市場における価値に着目するファンダメンタルズ分析ではない。

機械学習の分野において、株主価値に着目したファンダメンタルズ分析を行うモデルが提案されていない一方で、経済学的な分野では、一度株主価値を推定するファンダメンタルズ分析を用いた後に、株主価値を発行株式数で割ることで、株価を推定することが一般的である。

証券アナリストは、インタビューを通して対象企業や業界を分析することで、表

層的な財務情報だけでなく、詳細な商品や事業内容などの解像度の高い情報を元に、ファンダメンタルズ分析に基づいた経済学的な企業価値推定モデル(以下経済学的ファンダメンタルズ分析と呼ぶ)を用いて企業価値推定や投資価値推定を行っている。しかし、先述したFD規制によって、そうした解像度の高い情報の入手が難しくなり、人間による経済学的ファンダメンタルズ分析では予測が困難になってしまった [100]。

機械学習が飛躍的な発展を遂げ、企業の財務情報も取得することが可能になった今、市場における株主価値を考慮するファンダメンタルズ分析の機械学習モデルを用いることの価値は相対的に高まっており、今後も高まり続けると考えることができる。

そこで本研究では、既存のテクニカル分析の機械学習手法を用いるだけでなく、市場における価値に着目したファンダメンタルズ分析を機械学習を用いて行い、アンサンブル学習を用いた複合的な株主価値推定モデルを提案する。市場における株主価値を考慮するファンダメンタルズ分析の機械学習を実現するために、XBRL(eXtensible Business Reporting Language) ファイルを利用することによって、全ての上場企業の決算短信をはじめとした企業情報を取得するシステムを構築し、企業情報を一元化する。そうして、株価といった企業ごとにスケールが異なる値を株主価値に標準化し、推定対象とする。提案モデルの概要を図3.1に示す。

本研究では、2013年第1四半期から2014年第4四半期の8期において、日本の上場企業約3000企業、5年間分のデータを用いて学習を行い、分類・回帰予測の両方について、提案モデルを評価した。検証の結果、

- 経済学的ファンダメンタルズ分析に対して、提案モデルのファンダメンタルズ分析がより正確な推定を行えること
- テクニカル分析に加えてファンダメンタルズ分析を組み合わせた提案モデルが、テクニカル分析、さらにはファンダメンタルズ分析の各単体手法よりも高い精度の株主価値推定を行えること

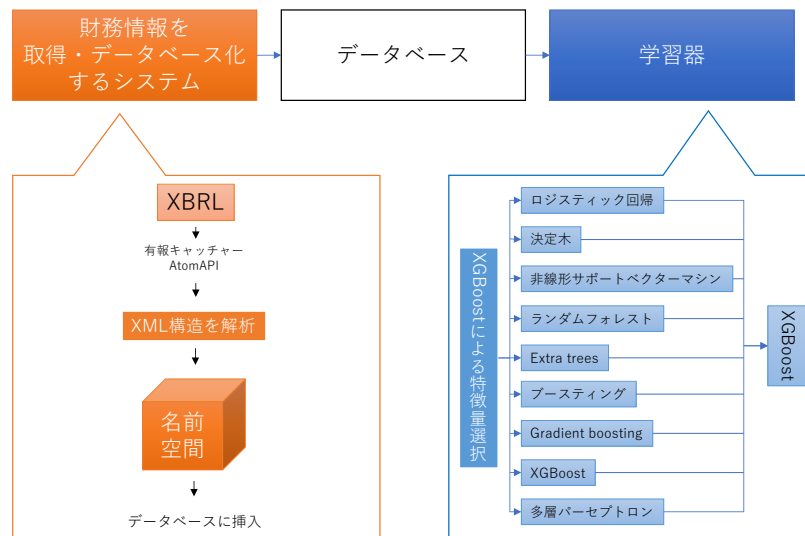


図 3.1: 深層学習を用いたアンサンブルモデルによる株主価値推定モデル

- XGBoost [101] の決定木の情報利得を利用した重要度分析により、予測対象が長期化するほど、ファンダメンタルズ分析が有効であることを

を示した。

3.2 提案手法

ファンダメンタルズ分析が中長期的な投資の際に重要であることはすでに 3.1 節で述べたが、人間による経済学的ファンダメンタルズ分析の問題として

- 人間が処理できるデータサンプル数の限界
- 人間の認知限界
- 複数の手法の組み合わせ方が体系立てられていないこと

があると洞察できる。

「人間が処理できるデータサンプル数の限界」及び「人間の認知限界」は、全企業の時系列データを用いて教師あり機械学習を行うことで、帰納的にパターンを見つけ出し、重要な素性を抽出することで解決できると期待する。XML の規格に

従って企業ごとの情報が記載されている XBRL ファイルが、ウェブ上に存在している。複数の企業の XBRL ファイルを取得し解析することにより、複数の企業情報を一元化することが可能である。複数の企業情報を一元化し、推定対象を株主価値にすることによって、株式市場における相対的な価値を考慮したようなファンダメンタルズ分析が、機械学習によって実現できる。

3つ目の「複数の手法の組み合わせ方が体系立てられていない」という問題に関しては、複数の分析手法を組み合わせるアンサンブル学習を用いて解決できると期待できる。

また、3.1節で述べた証券アナリストの行う企業価値推定や投資価値推定は、将来株価の上下の分類と、株価の適正価格や目標価格の回帰予測の2種類ある。よって本研究も、分類モデルと回帰モデルの2つを提案する。

3.2.1 概要

提案する「深層学習を用いたアンサンブルモデルによる株主価値推定モデル」の概要は図 3.1 に示した通り。既存の機械学習手法と同様に、テクニカル指標を入力素性としたテクニカル分析の機械学習モデル(以下「提案モデルのテクニカル分析」と呼ぶ)に加えて、XBRL ファイルから取得したファンダメンタルズ指標を用いて、市場における価値に着目したファンダメンタルズ分析の機械学習モデル(以下「提案モデルのファンダメンタルズ分析」と呼ぶ)がある。提案モデルは、提案モデルのテクニカル分析と提案モデルのファンダメンタルズ分析のモデル(以下ではまとめて「提案モデルの1層目」と呼ぶ)の予測結果を入力素性とするアンサンブル学習の非線形モデル(以下では「提案モデルの最終層」と呼ぶ)によって構成される。提案モデルの最終層は、最終予測が分類問題である場合は分類モデル、回帰問題である場合は回帰モデルを採用する。

複数の分析による予測を入力データとする非線形学習器を用いたアンサンブル

モデルによって、異なる分析手法を融合し、より予測性能が高いモデルを構築することを試みる本研究では、提案モデルの最終層に XGBoost を使用する。XGBoost を使用する理由は、

- 非線形モデルを使用することによって、分析間における非線形関係を表現できる
- 決定木を用いたアルゴリズムであるため、特徴量の重要度を情報利得から計算することができ、最終的にどの分析がどれだけ予測値に寄与しているかということが指標化できる

といったことがある。情報利得は、以下のように計算される。

$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right}) \quad (3.1)$$

ただし、 f は分割を行う特徴量であり、 I は不純度を表す関数、 D_p は親のデータセット、 D_{left} 及び D_{right} は子ノードのデータセットである。

3.2.2 企業情報の取得

企業のテクニカル指標は、多くの場合データベース化されており、Web から簡単に取得できる。一方で、企業のファンダメンタルズ情報は一般的に一元化されて公開されておらず、一元的にデータベース化するシステムを構築する必要がある。

XBRL ファイルとは、財務諸表などのビジネスレポートを XML の規格をベースとして電子化されたものである。本研究では、XBRL ファイルを利用し、すべての企業のビジネスレポートを一元的にデータベース化するシステムを構築した。

システム概要は以下の通り。

1. 適時開示情報閲覧サービス¹から企業のXBRLファイルをダウンロードする
2. 金融庁が公開しているタクソノミ要素リスト²を元に、各要素の名前空間を作成する
3. 作成した名前空間をもとに、XMLファイルを解析し、データベースにデータを挿入する

なお、本章で提案するモデルは、タクソノミ要素リストで定義されている要素のうち、決算短信、損益決算書、貸借対照表、キャッシュフロー表に含まれるものをすべてを入力素性の対象とし、テクニカル指標はトムソン・ロイターのものを用いた。

3.2.3 テクニカル分析のモデル

テクニカル分析においては、回帰モデルの場合、将来株主価値でなく一旦株価変化率を求める。企業ごとの株主価値は株価に比例しているため、対象企業の時系列データを学習に用いるテクニカル分析によって株価変化率を推定したのち、

$$\text{推定株主価値} = \text{現在株主価値} \times \text{推定株価変化率}$$

として株主価値を推定することができる。

今回、予測対象企業に対する市場の時系列相関を反映させるため、テクニカル指標をすべての企業、マーケット指数に対して計算し、すべての企業、マーケット指数のデータを合算したものを、企業 k の時刻 t における入力データ

$$\mathbf{X}_k^t = (\mathbf{x}_{c1}, \mathbf{x}_{c2}, \mathbf{x}_{c3}, \dots, \mathbf{x}_{cN}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iM})$$

とする。 x_{cj} は時刻 t における j ($j = 1, 2, 3, \dots, N$) 社のテクニカル指標ベクトル (始値, 高値, 安値, 終値, 出来高, 及び表 3.1 に記述するテクニカル指標からな

¹https://www.release.tdnet.info/inbs/I_main_00.html

²<https://www.fsa.go.jp/search/20200807.html>

るベクトル), x_{il} はマーケット指数 $l(l = 1, 2, 3, \dots, M)$ のテクニカル指標ベクトルである。また, 企業 k の時刻 t における正解データ y_k^t は予測対象企業 k の株価変化率とする。

テクニカル分析では長い系列長の時系列データを扱うため, 今回, 長期的な時系列データを扱うことが可能である Long Short Term Memory(以下, LSTM) [18] を学習器として用いる。また, 再帰的ニューラルネットワークの一つである LSTM を使用することによって, 学習の過程で過去の時系列相関を獲得できることが期待される。LSTM の再帰的な学習の過程

$$\mathbf{z}^t = \mathbf{f}(\mathbf{W}_{\text{in}}\mathbf{X}^t + \mathbf{W}_z\mathbf{z}^{t-1})$$

(\mathbf{f} は活性化関数, \mathbf{W}_{in} は入力層から隠れ層の重み行列, \mathbf{W} は隠れ層の重み行列, \mathbf{X}^t は時刻 t における入力データ, \mathbf{z}^t は時刻 t における隠れ層の出力である) によって獲得された重みは, 「対象企業の株価に過去の各情報がどれだけの影響を与えるか」という指標と捉えることができる。

一方, 経済学の分野では, 現代ポートフォリオ理論 [43] では株価の相関を表す値として

$$\beta = \frac{\frac{1}{n} \sum_{t=1}^T (\Delta A_t - \mu_{\Delta A})(\Delta \text{Index}_t - \mu_{\Delta \text{Index}})}{\sigma_{\Delta \text{Index}}}$$

が用いられているが ($\Delta A_t, \Delta \text{Index}_t$ は時刻 t における企業 A の株価 A 及び指数株価 $index$ の変化率, $\mu_{\Delta A}, \mu_{\Delta \text{Index}}$ は $t = 1\tilde{T}$ における平均値, $\sigma_{\Delta \text{Index}}$ は $t = 1\tilde{T}$ における標準偏差である), この数式モデルは時系列性を無視した単純な変化率の相関を導き出しているに過ぎない。再帰的に時系列を扱う LSTM を用いることで企業間の時系列の影響を学習し, より優れた予測を行えることが期待される。

3.2.4 ファンダメンタルズ分析のモデル

既存の株価予測研究のモデルの多くはファンダメンタルズ分析を取り入れたモデルではなかった。今回、全企業のファンダメンタルズ情報を所得し一元化したこと、及び予測対象を株主価値にすることによって、企業の市場における価値を考慮することが可能になる。

構築した企業情報をデータベース化するシステムによって、XBRLファイルを取得、解析することで企業活動情報をデータベース化する。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計算し、それに基づいて特徴量選択を行う。特徴量選択によって次元削減された特徴量を用いて複数の学習器(ロジスティック回帰, 決定木, 非線形サポートベクターマシン, バギング, ランダムフォレスト, Extra Trees, Adaboost, Gradient boosting, XGBoost, 多層パーセプトロン)を学習させる。各学習器の予測を入力データとして, XGBoostを学習・予測をするアンサンブルモデルを構築し, より予測性能が高いファンダメンタルズ分析のモデルの構築を試みる。

3.3 実験

3.3.1 実験目的

提案モデルに関して検証実験を行うことによって,

- 提案モデルのファンダメンタルズ分析が経済学的ファンダメンタルズ分析より高い精度の株主価値推定を行えること
- テクニカル分析に加えてファンダメンタルズ分析を組み合わせた提案モデルが, テクニカル分析, さらにはファンダメンタルズ分析の各単体手法よりも高い精度の株主価値推定を行えること

を示す。

3.3.2 実験する手法

本研究では，以下の手法の検証実験を行う．

手法 **EF** 既存の経済学的なファンダメンタルズ分析

手法 **MT** 提案モデルのテクニカル分析のモデル

手法 **MF** 提案モデルのファンダメンタルズ分析のモデル

手法 **MTMF** テクニカル分析とファンダメンタルズ分析のモデルをアンサンブルした提案モデル

EF

経済学的ファンダメンタルズ分析は，アセットアプローチ，マーケットアプローチ，インカムアプローチの3つがある．アセットアプローチでは，簿価純資産法を用いて検証する．今回は簡単のため，純資産額 = 推定株主価値とする．マーケットアプローチでは市場株価法を用いる．今回の検証実験においては，直近営業日60日間の終値平均を用いる．インカムアプローチでは，Discounted Cash Flow法（以下DCF法と呼ぶ）を用いる．DCF法は，将来キャッシュフローをリスク等を勘案した割引率によって現在価値に割り引いた金額を評価額とする評価方法で，ファイナンス理論に最も忠実であり実務的に様々な場面で利用されているものである．今回の検証実験では，四半期ごとのフリーキャッシュフロー（以下FCFと呼ぶ）に対して割引率を適用する．DCF法の具体的な方法は以下の通り．

1. Weighted Average Cost of Capital(負債の資本コストと株式の資本コストを負債比率と株式比率で加重平均したもの，以下WACCと呼ぶ)をCAPM [7]に基づいて計算
2. FCFの計算

$$FCF = \text{営業キャッシュフロー} - \text{投資キャッシュフロー}$$

3. FCF を WACC を用いて現在価値に割りもどし，定額モデルによって総和を「企業価値」とする

$$\text{企業価値} = \lim_{n \rightarrow \infty} \sum_{k=1}^n FCF \times \left(\frac{1}{1+WACC} \right)^{\frac{k}{4}}$$

4. 「企業価値」に非事業用資産を足したのち，有利子負債を差し引いたものを株主価値とする

$$\text{株主価値} = \text{企業価値} + \text{非事業用資産} - \text{有利子負債}$$

今回の検証実験において，近年のマイナス金利といった特殊な市場状態を考慮し，安定した予測を行うために Rd(負債コスト) 及び Rm(市場期待収益率) に対して，過去の数値の範囲内でパラメタチューニングを行い，仮想的に，Rd = 0.05, Rm = 0.1 とした．将来の FCF の予測が困難であることを仮想的に解決するために，本来は知ることのできない検証時点から将来 4 期分 FCF を既知とし，将来 5 期以降の FCF は検証時点から将来 4 期分 FCF の平均が継続すると仮定して，DCF 法を用いる．

経済学的ファンダメンタルズ分析の最終的な予測は，分類予測の場合，現在の株主価値との比較において単純平均が上であれば上昇予測，下であれば下降予測とする．回帰予測の場合，特にインカム分析の DCF 法では，無限等比数列を計算するため，割引率の計算や将来 FCF の予測値によって大きく値が変動してしまうことを懸念し，3つの手法の単純平均を予測値とするが，値が現在値から± 100%以上離れている場合は，予測値を± 100%に制限したものとする．

MT

検証の対象となる企業は，基本的に検証期間に上場している日本企業約 3000 企業である．しかし，計算時間の制約から，全期間において全企業の検証をすることは難しいことを考慮し，MT においては，企業ごとにモデルを学習・構築するの

ではなく、日経平均株価を構成する 225 銘柄 (2016 年 4 月時点) の検証期間から前 5 年間分の週足のデータを学習用データとした代替モデルを構築し、代替モデルを各企業に適用することで評価を行う。そして今回、週足データをデータセットの 1 タイムステップとし、LSTM の入力データを 30 ステップとする。

具体的に、データセットは、test dataset には検証時点から前 30 ステップのデータを用いる。学習用データは、test dataset に用いる 30 週間分のデータの直前から前 5 年間分遡ったデータを用いる。学習用データのうち前半 95% を train dataset、後半 5% を validation dataset に分割し学習を行う。なおデータセットを作成する時、予測を行うモデルを構築する際に未来情報を与えてはいけことに気をつける。

MF

MF を以下の手順で作成する。

step1 測対象期間の 1 期前から前 5 年間分 (20 期) の train dataset を用意する。

train dataset をさらにランダムサンプリングによって、1 層目の学習器の学習及びハイパーパラメタの最適化に用いる train first dataset と、2 層目の学習器 (XGBoost) の学習及びハイパーパラメタの最適化に用いる train second dataset に 2 等分割する

step2 train first dataset を用いてハイパーパラメタの調整を行う

step3 train first dataset と調整したハイパーパラメタを用いて 1 層目の学習器層目の学習器 (ロジスティック回帰, 決定木, 非線形サポートベクターマシン, バギング, ランダムフォレスト, Extra Trees, Adaboost, Gradient boosting, XGBoost, 多層パーセプトロン) を作成する

step4 作成した一層目の学習器を用いて、train second dataset による予測をそれぞれの学習器ごとに出力する。これらの出力を予測データセットとする

step5 出力した予測データセットを用いて2層目のXGBoostのハイパーパラメタの調整を行う

step6 調整したハイパーパラメタと予測データセットを用いて2層目のXGBoostを学習させる

MTMF

MTと、MFを用いて、MTMFの最終層のXGBoostのパラメタ最適化及び学習を行う。

step1 検証期間における、2013年第1四半期から2014年第4四半期のMT及びMFの分類、回帰予測をそれぞれ行い、3.1節で述べたように入力用の予測データセットを作成する

step2 予測データセットのうち、予測対象期間のものをtest dataset、残りをtrain datasetとするtrain datasetをシャッフルし、データセットを5分割する。ハイパーパラメタサーチを行い、最終層のXGBoostのハイパーパラメタを最適化する

step3 最適化されたハイパーパラメタとtest datasetを用いて最終層のXGBoostを学習する

3.3.3 検証実験の設定

評価方法

検証実験としての公平性を保つために、検証する期間は、2013年第1四半期から2014年第4四半期までの8四半期とする。また、ファンダメンタルズ分析とテクニカル分析の優位性が予測期間の長さによって異なる可能性を考慮し、異なる予測期間に対して検証を行う。証券アナリストが、3ヶ月後、1年後の投資価値評価を行うことは一般的であり、特に、決算短信が出てから次の決算短信が出るまで約60営業日(約3ヶ月)あることを踏まえて、予測対象期間を60営業日、240営業日とする。

業日(約1年)の2つについて検証する。

ファンダメンタルズ分析の株主価値の推定に対する寄与度を正確に評価し、できるだけ市場のランダム要素を排除したい。そこで、各企業の決算短信が発表された翌日からの、将来の平均の株主価値を推定し、分類予測の場合、株主価値が予測時点から上昇するか下降するかの正解率によって、回帰予測の場合、対象企業における株主価値の平均誤差率によって評価する。

データセット

企業ごとのテクニカル指標に関して、始値、高値、安値、終値、出来高はトムソン・ロイターのデータを用いる。さらに表3.1に記述するテクニカル指標を計算して追加する。

表 3.1: 採用するテクニカル指標一覧

テクニカル指標	式	概要
$ROC_0(c, x)$	$\frac{x_t}{ct} \quad x_t = [s_t, h_t, l_t]$	終値に対する始値、高値、安値の割合
$ROC_k(c)$	$\frac{c_t}{c_{t-k}} \quad k = 1, 5, 20, 60, 120, 240, 360, 720$	kステップ前からの終値変化率
$ROC_k(v)$	$\frac{v_t}{v_{t-k}} \quad k = 5, 20, 60, 120, 360$	kステップ前からの出来高変化率
$Disparity_k(c)$	$\frac{c_t}{\frac{1}{k} \sum_{i=t-k+1}^t c_i} \quad k = 5, 25, 75$	終値の k ステップ移動平均乖離率
$RSI_k(c)$	$1 - \frac{1}{1 + \left(\frac{\sum_{i=t-k+1}^t U_{t-i}}{\sum_{i=t-k+1}^t D_{t-i}} \right)}$ <p>where $U_t = \begin{cases} 1, & c_t - c_{t-1} > 0 \\ 0, & c_t - c_{t-1} \leq 0 \end{cases}$</p> <p>where $D_t = \begin{cases} 1, & c_t - c_{t-1} < 0 \\ 0, & c_t - c_{t-1} \geq 0 \end{cases}$</p>	kステップで値上がりしたステップの割合 $k = 14, 20, 60$
$GC_{k,l}(c)$	$\begin{cases} 1, & a_t(k) - a_t(l) > 0 \ \& \ a_{t-1}(k) - a_{t-1}(l) \geq 0 \\ 0, & otherwise \end{cases}$ <p>where $k < l, \quad a_t(k) = \frac{1}{k} \sum_{i=t-k+1}^t c_i$</p>	ゴールデクロス $(k, l) = (5, 25), (25, 75)$
$DC_{k,l}(c)$	$\begin{cases} 1, & a_t(k) - a_t(l) < 0 \ \& \ a_{t-1}(k) - a_{t-1}(l) \leq 0 \\ 0, & otherwise \end{cases}$ <p>where $k < l, \quad a_t(k) = \frac{1}{k} \sum_{i=t-k+1}^t c_i$</p>	デットクロス $(k, l) = (5, 25), (25, 75)$

MTにおいては、日経平均株価を構成する225銘柄(2016年4月時点)の検証時点から前5年間分の正規化された週足のデータ、約56000データサンプルを用いて学習する。

企業ごとのファンダメンタルズ指標は、XBRLファイルを解析することで得られた決算短信、損益決算書、貸借対照表、キャッシュフロー表のデータを用いる。さらに、Yahoo!ファイナンス³から、業種分類、設立年月日、市場名、上場年月日、従業員数(単独)、従業員数(連結)、単元株数、平均年齢、平均年収といった企業情報を参照、特徴量として追加する。なお、上場年月日、設立記念日などは現在時刻との差分を計算することによって経過時間を割り出し、特徴量として追加する。そうして、取得したデータを用いて表3.2に記述するファンダメンタルズ指標を計算・追加する。本検証実験では、前期のデータを参照する欠損値補完、外れ値の置換、標準化をおこなう。また、取得したデータのうち、損益決算書、キャッシュフロー表、及び決算短信の一部に含まれるデータは、四半期ごとの独立したデータでなく、年度内の累計データである。そこで、四半期ごとのデータを並列して対等に扱うため、累計値から四半期ごとの値を逆算する形で求め、修正する。MFでは、そうして作成される約3000企業の5年間分の四半期データ、約60000データサンプルを学習に用いる。

³<https://finance.yahoo.co.jp/>

表 3.2: 採用するファンダメンタルズ指標一覧

ファンダメンタルズ指標	式	概要
<i>PER</i>	$\frac{\text{時価総額}}{\text{当期純利益}}$	株価収益率
<i>EPS</i>	$\frac{\text{純利益}}{\text{発行済株式総数}}$	一株あたりの利益
<i>BPS</i>	$\frac{\text{株主資本}}{\text{発行済株式総数}}$	一株あたりの株主資本
<i>PBR</i>	$\frac{\text{現在株価}}{BPS}$	株価純資産倍率
配当利回り	$\frac{\text{一株あたりの配当}}{\text{株価}} \times 100$	株価に対する年間配当金の割合を示す指標
<i>ROE</i>	$\frac{EPS}{BPS}$	自己資本利益率
<i>ROA</i>	$\frac{\text{純利益}}{\text{総資産}}$	総資産利益率
自己資本比率	$\frac{\text{総資産} - \text{負債}}{\text{総資産}} \times 100$	自己資本が総資産を占める割合
有利子負債自己資本比率	$\frac{\text{有利子負債}}{\text{自己資本}} \times 100$	有利子負債が自己資本を占める割合
<i>EV</i>	時価総額 + 有利子負債 - 貯金	貸貸借表から導かれる企業価値
<i>EBITDA</i>	営業利益 + 減価償却費	利払い前・税引き前・減価償却前・その他償却前利益
<i>EV/EBITDA</i>	$\frac{EV}{EBITDA} \times 100$	簡易買収倍率、企業の割安性を測る指標
流動比率	$\frac{\text{流動資産}}{\text{流動負債}} \times 100$	企業の短期的な支払能力を判断する指標
当座比率	$\frac{\text{当座資産}}{\text{流動負債}} \times 100$	企業の短期の負債に対する支払い能力を判断する指標

ハイパーパラメタの取り扱い方

本研究では、ハイパーオプト [102] を用いてハイパーパラメタサーチを行なう。なお、MTにおけるLSTMは計算コストが高いため、ハイパーオプトは行わず、隠れ層のノード: 200, ドロップアウト: 0.15, 出力層の活性化関数: [分類問題: softmax, 回帰問題: linear], 学習率: 0.00001, 最適化法: RMSprop [103], 損失関数: [分類問題: categorical cross entropy, 回帰問題: mean squared error] とハイパーパラメタや最適化法を設定した。

3.3.4 実験結果

検証実験による結果を表 1, 表 2 に示す。

表 3.3: 分類予測における正解率の比較 (%)

検証期間	予測対象	EF	MT	MF	MTMF
2013Q1	60日	50.46	48.35	53.02	58.43
	240日	46.57	38.3	69.63	72.76
2013Q2	60日	45.51	55.72	59.65	62.82
	240日	54.43	63.57	67.92	71.75
2013Q3	60日	56.43	54.84	47.8	47.8
	240日	52.37	50.17	68.96	76.36
2013Q4	60日	42.92	73.13	63.54	51.84
	240日	60.58	61.55	74.7	78.5
2014Q1	60日	52.23	57.11	55.95	56.4
	240日	53.88	48.7	70.59	69.9
2014Q2	60日	40.51	35.6	64.66	60.83
	240日	53.04	59.63	74.21	73.59
2014Q3	60日	34.97	39.95	67.88	62.49
	240日	52.61	54.54	65.44	66.15
2014Q4	60日	52.28	45.35	59.56	59.32
	240日	54.23	46.05	53.35	54.37

表 3.4: 回帰予測における正解率の比較 (%)

検証期間	予測対象	EF	MT	MF	MTMF
2013Q1	60日	29.96	12.44	14.52	10.25
	240日	32.39	38.8	19.31	18.44
2013Q2	60日	30.32	11.54	11.21	8.14
	240日	31.65	20.58	15.25	14.73
2013Q3	60日	26.92	9.1	25.55	10.09
	240日	30.3	20.91	22.5	17.03
2013Q4	60日	28.5	12.63	11.84	9.06
	240日	29.54	25.7	16	14.77
2014Q1	60日	28.7	8.24	9.64	8.23
	240日	29.16	16.83	15.36	13.71
2014Q2	60日	28.83	7.9	8.73	7.46
	240日	27.8	15.64	13.25	12.76
2014Q3	60日	32.53	10.31	9.07	7.55
	240日	29.18	14.11	13.48	12.86
2014Q4	60日	28.06	8.36	8.55	7.59
	240日	28.46	14.88	16.32	15.02

表 3.3, 表 3.4 は, 検証期間が 2013 年第 1 四半期から 2014 年第 4 四半期までの 8 四半期における, 推定対象を営業 60 日平均の株主価値及び営業 240 日平均の株主価値とした場合のそれぞれの評価結果であり, 表 3.3 は各モデルの株主価値の上下予測の正解率, 表 3.4 は株主価値の平均誤差率を表している. 表 3.3 分類予測における正解率の比較において, EF と MF の正解率の評価結果を比較すると, 16 検証パターン中 14 パターンで MF の正解率が優れている. 同様に, 表 3.4 の回帰予測における平均誤差率の比較において, 全ての検証パターンで MF の平均誤差率が優れている.

MT と MTMF を比較すると, 分類問題では 16 検証パターンで 13 パターンにおいて MT よりも MTMF が高い正解率を示していて, 回帰問題では 16 検証パター

ン中14パターンでMTよりもMTMFが優れた平均誤差率を示している。MTの正解率平均は52.04%であるのに対し、MTMFの正解率平均は63.96%であり、11.92ポイント優れている。MTの誤差率平均は15.50%であるのに対し、MTMFの誤差率平均は11.73%であり、3.77ポイント優れている。

MTとMFの双方と、MTMFを比較すると、分類問題では16検証パターン中8パターンにおいてMT及びMFよりもMTMFが高い正解率を示していて、回帰問題では16検証パターン中14パターンでMT及びMFよりもMTMFが優れた平均誤差率を示している。特に、全検証パターンの平均を取ると、分類予測の場合、提案モデルを構成する各分析の正解率平均が57.79%であるのに対し、それらをアンサンブルした提案モデルは63.96%であり、6.16ポイント優れている。回帰予測の場合、提案モデルを構成する各分析の平均が14.95%であるのに対し、それらをアンサンブルした提案モデルは11.73%であり、3.22ポイント優れている。

表 3.5: 2014 年第 4 四半期・営業 240 日平均の分類予測におけるファンダメンタルズ指標
重要度

順位	項目	順位	項目
1	株価	26	たな卸資産の増減額
2	PBR	27	前年比たな卸資産の増減額
3	配当利回り	28	売上債権の増減額
4	前年比その他有価証券評価差額金	29	前年比経常利益又は経常損失
5	前年比受取配当金	30	未払消費税等の増減額
6	前年比貸倒引当金	31	前年比固定資産売却益
7	その他有価証券評価差額金	32	特別利益
8	前年比繰延税金資産	33	その他費用
9	前年比為替換算調整勘定	34	前年比財務活動によるキャッシュ・フロー
10	自己資本比率	35	前年比その他の包括利益
11	EBITDA	36	為替差損益
12	前年比為替換算調整勘定	37	前年比現金及び預金
13	特別損失	38	前年比その他資産
14	前年比包括利益	39	前年比支払利息
15	PER	40	前年比特別損失
16	貸倒引当金	41	前年比退職給付に係る負債
17	前年比売上総利益又は売上総損失	42	賞与引当金の増減額
18	前年比受取手形及び売掛金	43	発行株式数
19	前年比流動資産	44	前年比貸倒引当金
20	前年比負債	45	法人税等
21	前年比セール・アンド・リースバックによる収入	46	売上債権の増減額
22	前年比退職給付に係る負債の増減額	47	前年比その他費用
23	前年比利益剰余金	48	前年比小計
24	EV	49	貸倒引当金
25	従業員数	50	建物

表 3.6: 2014 年第 4 四半期・営業 240 日平均の回帰予測におけるファンダメンタルズ指標重要度

順位	項目	順位	項目
1	EV	26	流動資産
2	連結総従業員数	27	その他の営業収入
3	株価	28	自己株式
4	株式発行数	29	有利子負債自己資本比率
5	PBR	30	有価証券
6	PER	31	持分法による投資損益
7	従業員数	32	連結の範囲の変更を伴う子会社株式の売却による収入
8	配当利回り	33	営業利益又は営業損失
9	法人税等の支払額	34	建設仮勘定
10	配当金の支払額	35	為替予約
11	資産	36	前年比その他収入
12	資本剰余金	37	前年比為替換算調整勘定
13	不動産の企業	38	新規連結に伴う現金及び現金同等物の増加額
14	株主資本	39	未払法人税等
15	BPS	40	その他権利
16	利益剰余金	41	EV/EBITDA
17	短期借入金	42	繰延税金負債
18	貸倒引当金	43	有形固定資産の売却による収入
19	設立からの経過時間	44	当期純利益又は当期純損失
20	流動比率	45	法人税等調整額
21	前年比その他の包括利益	46	その他の資産・負債の増減額
22	税引前当期純利益又は税引前当期純損失	47	定期預金の増減額
23	前年比為替差益	48	保険積立金
24	長期借入金	49	為替換算調整勘定
25	支払利息	50	減価償却費

表 3.5, 表 3.6 は, MF の検証過程で, XGBoost に基づいて特徴量ごとに情報利得の平均を計算し, 順位付けを行ったものである. 推定時点での株主価値に直接比例する EV(Enterprise Value) や株価, 株主発行数が上位に来ているだけでなく, PER(株価収益率) や PBR(株価純資産倍率) といった一般的なファンダメンタルズ指標も上位にきている. また, 従業員数や法人税等の支払額など, 直接株主資本

に関連しないような項目も重要な特徴量として抽出されている。このように、MFは、ファンダメンタルズ指標から株主価値推定に重要な特徴量を抽出していることがわかる。

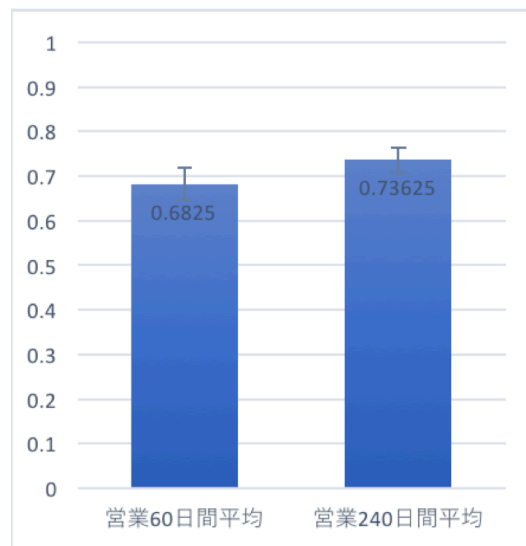


図 3.2: 分類予測におけるファンダメンタルズ分析の重要度の割合

また、ファンダメンタルズ分析の最終的な予測における重要度 (最終的な推定結果に対する寄与度) が全体に占める割合を図 3.2, 図 3.3 に示す。図 3.2, 図 3.3 は、分類予測及び回帰予測において、営業 60 日間平均の予測に対して営業 240 日間平均の予測の方がよりファンダメンタルズ分析の寄与度が有意に大きいことを示している。

3.4 実験に対する考察

実験で行った EF と MF の評価結果から、経済学的なファンダメンタルズ分析と比較して、提案モデルのファンダメンタルズ分析がより正確な株主価値推定を行えることがわかった。機械学習によって、約 60000 サンプルのデータセットから帰納的な推定を行ったこと、そして図 3.2, 図 3.3 のように重要な特徴量を 1000 項

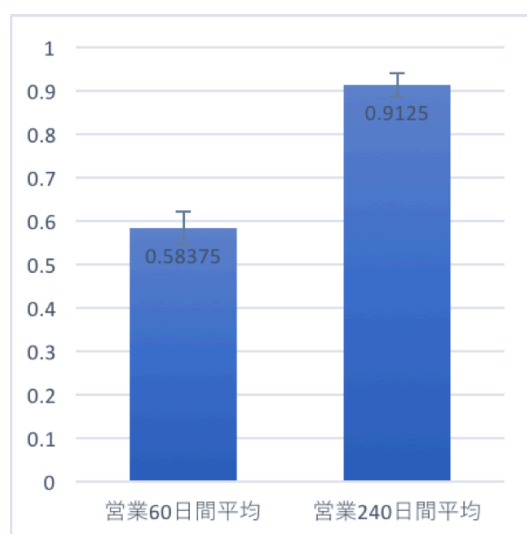


図 3.3: 回帰予測におけるファンダメンタルズ分析の重要度の割合

目ほどの企業情報の特徴量から抽出することによって、人間の処理できるデータサンプル数の限界及び人間の認知限界の問題を解消し、より精度の高い株主価値推定が行えたと考えられる。

また、表 3.5、表 3.6 において、キャッシュフローに関連する項目が株主価値推定における重要な特徴量の上位にきていないことは興味深い。なぜなら、代表的な経済学的なファンダメンタルズ分析である DCF 法は、企業価値は将来生み出すキャッシュフローの割引現在価値によって決定するというファイナンス理論に基づいており、キャッシュフローに関連する項目が重要な特徴量として抽出されることが期待できたからである。ファイナンス理論と、実際に機械学習を用いて解析的に得られた結果が必ずしも一致しないと考えられる。

法人税等の支払額が重要な特徴量として抽出されていることにも注目すべきである。ファイナンスの理論で企業価値の推定の際にキャッシュフローに着目するのは、期間業績の適正化から財務会計の基準が企業ごとに異なり、客観的な比較が正確に行えないという理由がある。法人税の計算は法人税等の規定によって会計計算とは切り離されて別個に行われるものであり、法人税は利益に対して計算さ

れるため、法人税等の支払額は実際の利益を比較可能な形で反映していると考察できる。このように、提案モデルによって既存の経済学的なファンダメンタルズ分析では捉えていなかった項目や指標を抽出し、既存手法では着目していなかった指標や観点から企業価値を評価することが可能になると期待できる。

MTとMTMFの比較から、テクニカル分析に加えてファンダメンタルズ分析を組み合わせた提案モデルが、テクニカル分析に対して優れた推定を行えることがわかる。また、図3.2、図3.3の結果から、予測対象が長期化するほど、ファンダメンタルズ分析が有効であると推論できる。よって今回のような中長期的な投資の際に、テクニカル分析に加えて、長期的である場合により重要度が高まる株主価値に着目したファンダメンタルズ分析を行うことによって、より優れた予測を行えると考察する。

さらに、MTとMF、MTMFの評価結果の比較から、ファンダメンタルズ分析とテクニカル分析を融合させた提案モデルが、単一の分析手法より汎化性能が高いモデルとなることが示された。

「テクニカル分析は市場の動きの研究に専念するのに対し、ファンダメンタルズ分析は価格を上下させたり、あるいは同じレベルにとどめさせたりする原因となる、需要と供給関係に焦点をあてる」[43]とあるように、ファンダメンタルズ因子とテクニカル因子は、それぞれ価格変動の原因と観測結果に対応している。ファンダメンタルズ因子とテクニカル因子という金融データの関係性の特徴として、ノイズトレーダー [43] の存在ゆえ、原因と観測結果が非合理である場合が存在する時系列データであることが挙げられる。例えば、ある銘柄Aの株価が下落し、その銘柄Aのファンダメンタルズ因子から導かれる株価に対して割安になったとする。割安であるならば、その銘柄Aを買うことが合理的な行動であり、銘柄Aの株価は上昇すると期待されるが、株価が下落したことにより、ノイズトレーダーは悲観的な評価を行って銘柄Aを売却するといった非合理的な行動を取り、結果として銘柄Aの株価は一定、またはさらに下落する可能性がある。

本研究は、より一般的に、原因と観測結果に非合理的な関係が存在している時系列のデータに対して、それぞれ原因に着目したモデルと観測結果に着目したモデルを独立に構築し、アンサンブルモデルを用いることで、予測精度を向上することができ、原因と観測結果のデータが予測結果にもたらす影響を定量的に分析できる可能性を示唆している。

例えば、インターネット広告の将来 cost per action(以下 CPA と呼ぶ)を予測する際に、インターネット広告の素性の特徴量(広告の商品, 色, 予算, 入札額, ターゲット層など)と、インターネット広告効果の観測結果の特徴量(インプレッション数やクリック数, CPA など)を用いる場合が考えられる。広告の素性は広告効果の観測結果に影響を与えるという因果関係が存在し、そして広告効果の観測結果は人の気まぐれや誤クリックといった合理的でない行動による結果を含んでいる。このような場合に、本研究の提案モデルのように、広告の素性を入力データとする提案モデルのファンダメンタルズ分析のモデルと、広告効果の観測結果を入力データとする提案モデルのテクニカル分析のモデルをそれぞれ独立に構築し、アンサンブルモデルを用いることで、予測精度を向上させ、広告の素性と観測結果を切り分けて予測結果に対する定量的な分析を行える可能性がある。

また本研究の提案モデルの適用可能性の他の例として、インターネット通販サイトにおける商品の売上予測を行う場合が考えられる。例えば商品 A と商品 B の2つの商品があり、値段以外の素性は全て同じで、商品 A の方が値段が安いとする。この場合、商品 A を購入することが合理的な行動であるが、知人が商品 B を購入していたなどの理由から、商品 B が商品 A に対して多く売れていると消費者が感じた場合、より多く売れている商品を購入したいという消費者の心理が働き、実際には商品 A よりも商品 B がさらに売れるという可能性が考えられる。商品の値段や種類、機能などといった素性のデータを特徴量とする提案モデルのファンダメンタルズ分析のモデルと、インターネット通販サイトにおけるインプレッションやクリック数、販売個数などの観測結果の時系列データを入力データとする提

案モデルのテクニカル分析のモデルをアンサンブルするというように、本研究の提案モデルを適用することで予測精度を向上させ、商品の素性と、観測結果がそれぞれ予測結果に与える影響を定量的に分析できる可能性がある。

今回の検証実験では、60日、240日と2つの期間についての推定を行った。図3.2、図3.3から、予測期間が異なれば予測に対する手法の寄与度も異なることがわかったが、十分な学習データが存在していれば、提案モデルは期間に応じて適化された分析手法の組み合わせを実現するため、より短い期間またはより長い期間にも適応することができると思う。また、今回の検証実験は日本の株式市場に対して行ったが、企業情報を取得するシステムを変更する必要があるが、十分な学習データさえ存在すれば他国の市場に適用することが可能であるはずである。

複数の手法を組み合わせるアンサンブルモデルにより、より予測性能が高いモデルを実現する提案モデルは、異なる分析手法を追加できる拡張性を持っており、さらに予測性能が高いモデルを実現できる可能性を持っている。例えば今回、ファンダメンタルズ分析として主にファンダメンタルズ指標の数値情報を特徴量としているが、実際の株式分析においては、決算短信の文字情報などを参考にすることが普通である。提案モデルは、事業内容や将来の方針などの文字情報から予測する分析を加えることが可能であり、より予測性能を上げられる可能性を持っている。

3.5 おわりに

本章では、既存のテクニカル分析の手法を用いるだけでなく、市場における価値に着目したファンダメンタルズ分析を機械学習を用いて行い、アンサンブル学習を用いた複合的な株主価値推定モデルを提案した。

XBRL ファイルを利用し、全ての上場企業の企業情報を取得するシステムを用

いて企業情報を一元化した。推定対象を標準化された株主価値にすることで、機械学習によるファンダメンタルズ分析を可能にした。機械学習によって、約60000サンプルのデータセットから帰納的な推定を行ったこと、そして表3.5のように重要な特徴量を1000項目ほどの企業情報の特徴量から抽出することによって、人間の処理できるデータサンプル数の限界及び人間の認知限界の問題を解消し、より精度の高い株主価値推定が行えることを示した。

そして、アンサンブルモデルによって異なる分析を融合させた提案モデルが、単一分析手法より予測性能が高いモデルとなることを示した。また、提案モデルによって今まで独立で研究されていた複数の既存手法を予測期間・対象に応じて組み合わせることが可能となり、さらに優れた予測を行えるようになる可能性について述べた。

第4章 企業の事業内容に着目した共分散行列 による深層学習ポートフォリオ最適化

インターネットやAIといった技術革新や、COVID-19による社会環境の変化に伴って、企業のビジネスモデルや事業内容が多様化している。本章では、株価情報だけを考慮するのではなく、ニューラルネットワークを用いて企業のビジネスモデルや事業内容のベクトル表現を獲得し、ポートフォリオの共分散行列を予測する手法を提案する。

4.1 はじめに

金融分野における取引戦略において、リスクを最小化しながらリターンを最大化するポートフォリオの構築は、学術的にも実務的にも重要な研究分野である。リターンの分布の性質に基づいて最適ポートフォリオを構築する平均分散アプローチ [43] は、古典的なポートフォリオ最適化手法であり、理論的にも実務的にも広く応用されてきた。平均分散アプローチの枠組みでより優れたポートフォリオを構築するには、i) ポートフォリオの将来の期待収益率をより精緻に推定する。ii) ポートフォリオの将来の分散をより精緻に推定する。という2つのアプローチが存在する。

本研究は、より優れた株式ポートフォリオの構築のために、ポートフォリオの将来の分散をより精緻に予測することを目的とする。将来のポートフォリオの分散は、ポートフォリオを構成する資産の分散・共分散を要素に持つ共分散行列で表現される。本研究は、株式ポートフォリオの共分散行列を精緻に予測する手法を提案する。

過去の株価情報から計算された標本共分散行列を、将来のポートフォリオの共分散行列として予測することは、基本的なアプローチである一方で、以下のような課題が指摘されている [38, 104].

1. 株価の非線形なダイナミクスを捉えることができないこと.
2. ポートフォリオの構成する資産の数に対して、時系列のサンプル数が不十分になる. 低ボラティリティーの固有要因のリスクを過小に評価し、高ボラティリティーの固有要因のリスクを過大に評価する傾向があること.
3. 時系列普遍性を前提にしているが、実際はその限りではないこと.

そこで、[39]では、共分散行列と、 β を単一指標として用いた共分散行列を加重平均することで、推定するパラメタ数に対して時系列サンプリング数が不足するという課題に対処する手法を提案した. [40]では、ファクターモデルの残差項目に着目し、スパースな分散・共分散行列を導入することで、より精緻なポートフォリオの分散・共分散を表現している. 他にも、平均分散アプローチ [43]の枠組みを超えて、収益率の期待値を用いずに、分散のみを用いてポートフォリオを構築するリスクベースのポートフォリオの研究もなされている [44-48].

既存研究の多くは、過去の株価情報をはじめとする定量的に観測可能なデータから、数学的・統計的手法を用いてポートフォリオの分散・共分散を推定する手法を提案し、より優れたポートフォリオの構築可能性を広げた. 一方で、企業のビジネスモデルや事業内容といった企業そのものが持つ定性的なデータを考慮して、株式ポートフォリオの分散・共分散を推定する手法は数少ない.

ビジネスモデルや事業内容が多様化し、SARS-CoV-2(以下、COVID-19) [105]による社会環境の変化が進む中で、企業の定性的な情報を考慮して株式ポートフォリオを構築する重要性は高まっている. 近年、Software as a Service(以下、SaaS)という新しいビジネスモデルを選択する企業が存在している. 従来の売り切り型のモデルと違って、SaaSはソフトウェアを利用する権利をサブスクリプション型で提供するビジネスモデルだ. 企業の成長に伴って、多角化を進める企業も存在

する。例えば、ソニー株式会社(6758)は、「電気機器」に業種分類されるが、ゲームや音楽、金融などの6つの事業セグメントを有しており、事業ポートフォリオを構築している。ビジネスモデルや事業内容に応じたリスクの要因や程度が存在し、ビジネスモデルや事業内容を考慮してポートフォリオの共分散行列を予測することが重要と考えられる。

ポートフォリオの共分散行列を予測する場合、ポートフォリオを構成する資産のリスク、及び資産間のリスクの相関が精緻に予測できればよい。定量的に観測可能な株価データに限らず、ビジネスモデルや事業内容などの企業の性質に着目することで、より精緻にポートフォリオの共分散行列を予測できるという仮説を立てる。

ポートフォリオの共分散行列を予測する際に、企業のビジネスモデルや事業内容といった定性的なデータを利用することは、データが非構造的であるため難しかった。一般的に、企業のビジネスモデルや事業内容は、投資家向け資料として、決算短信や決算説明資料に記述されているが、そうしたデータを一元化し必要な情報を抽出する必要がある。予めヒトが定義した業界分類の情報を用いることはできても、複数の事業ポートフォリオを持つ企業や、多種多様なビジネスモデルの全てを体系的に取り扱うことは、ヒトの認知処理限界故に難しい。

そこで本研究では、投資家向け資料からビジネスモデルや事業内容といった定性的な情報を抽出し、事業のベクトル表現を用いて共分散行列を予測する機械学習手法を提案する。過去の株価情報をはじめとする定量的に観測可能なデータから数学的・統計的手法を用いて導出される共分散行列だけでなく、ビジネスモデルや事業内容を反映した共分散行列もポートフォリオ最適化に有用であると期待される。

実験では、ビジネスモデルや事業内容を表すベクトル表現を用いて共分散行列を導出する提案手法の予測精度を比較検証した。日本株式市場を対象に、2016年

から2018年までの情報を用いて共分散行列の予測ならびにポートフォリオの構築を行い、2019年から2020年の期間で共分散行列の予測精度の評価とポートフォリオのパフォーマンスの検証を行った。その後、ビジネスモデルや事業内容を表すベクトル表現を用いて、企業の関連性やリスクの分析を行った。

本研究の貢献は、以下の通りである。

- 投資家向け資料からビジネスモデルや事業内容といった定性的な情報を抽出し、事業のベクトル表現を用いて共分散行列を予測する機械学習手法を提案し、より精緻な共分散行列の予測を可能にしたこと。
- ビジネスモデルや事業内容を反映した共分散行列が、株式ポートフォリオの最適化に有用であることを示したこと。
- 提案手法によって獲得できる事業内容を表すベクトル表現を用いることで、ビジネスモデルの多様化やCOVID-19の社会環境の変化に応じたリスクの分析可能性を高めたこと。

4.2 提案手法

本章では、図4.1で示すように、投資家向け資料からビジネスモデルや事業内容といった定性的な情報を抽出し、ニューラルネットワークによる事業のベクトル表現を用いた共分散行列、Deep-Business Covariance Matrix(以下、Deep-BCMと呼ぶ)を導出する機械学習手法を説明する。

4.2.1 決算短信及び有価証券報告書のデータベース化とテキストの取得

ビジネスモデルや事業内容といった企業そのものが持つ定性的なデータは、決算短信や有価証券報告書といった投資家向け資料にテキスト情報として記述されていることが一般的であり、定量的に評価・一元管理されていることは少ない。そのため、ビジネスモデルや事業内容といった企業の性質を表すデータを用いた手

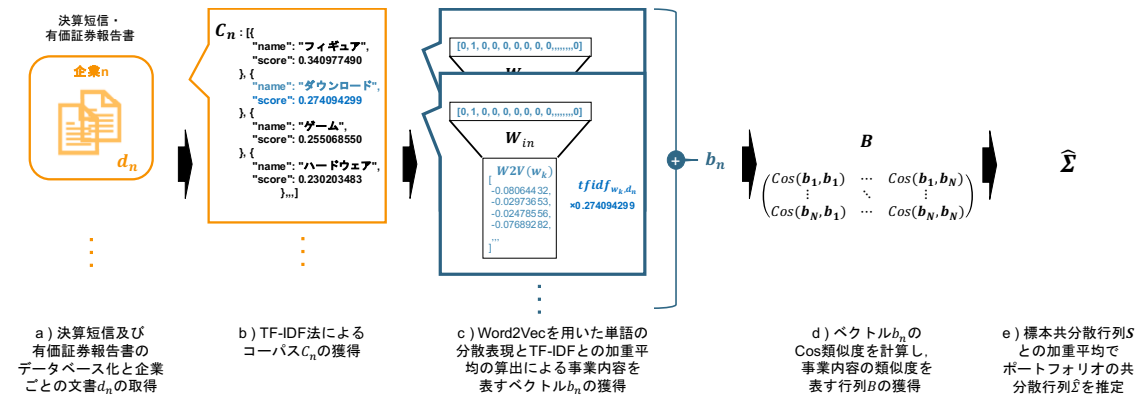


図 4.1: 提案する企業の事業内容に着目した共分散行列の獲得手法の概要

法の実現には、投資家向け資料のデータベース化と、テキストの取得を実現する必要がある。

本研究では、決算短信及び有価証券報告書を対象とし、財務諸表をはじめとする数値情報が記述されている XBRL¹ファイルと、事業内容や事業概要が書かれている PDF ファイルを取得し、データベース化する。

データ取得の流れは以下の通り。

1. EDINET²及び適時開示情報閲覧サービス³から企業の XBRL ファイル及び PDF をダウンロードする。
2. 金融庁が公開しているタクソノミ要素リスト⁴を元に、各要素の名前空間を作成する。
3. 作成した名前空間をもとに、XML ファイルを解析し、データベースにデータを挿入する。
4. 取得した PDF から PDFMiner⁵を用いてテキスト情報を抽出し、データベ

¹eXtensible Business Reporting Language の略。財務諸表などのビジネスレポートを XML の企画をベースとして電子書籍化されたもの

²<https://disclosure.edinet-fsa.go.jp/>

³https://www.release.tdnet.info/inbs/I_main_00.html

⁴<https://www.fsa.go.jp/search/20200807.html>

⁵<https://pypi.org/project/pdfminer/>

スにデータを挿入する。

4.2.2 ニューラルネットワークを用いた事業内容のベクトル表現の獲得

企業ごとに4.2.1項で取得したビジネスモデルや事業内容が記述されている文章を構成する特徴的な単語をTF-IDFで抽出して重み付けを行い、単語分散表現の加重平均を算出することで事業のベクトル表現を獲得する。

本研究の目的は、投資家向け資料からビジネスモデルや事業内容といった定性的な情報を抽出し、事業のベクトル表現を用いて共分散行列を精緻に予測することである。その上で、ビジネスモデルや事業内容が記述されている文章から事業のベクトル表現を獲得する場合、売上高や利益などの財務情報や、投資家向け資料に形式的に記述される事業内容以外の情報を解析対象から外すことによって、企業が固有にもつビジネスモデルや事業内容に注目したい。事業内容が記述されている文章からベクトル表現を獲得する際、Doc2Vec [106]の手法やSentenceBERT [107]によって文書全体の情報を活用することも検討できるが、ビジネスモデルや事業内容以外のテキスト情報も多く含む決算短信及び有価証券報告書では、ビジネスモデルや事業内容が記述されている文章を構成する特徴的な単語を重み付けして抽出し、単語の分散表現の平均を算出する手法が望ましいと考えられる。実際に、文書ベクトルを獲得する際に、Doc2Vec [106]よりも単語分散表現の平均を用いたシンプルな手法の方が実タスクによっては精度が高いとする研究 [108]もある。

まず、企業 n の特定期間の決算短信及び有価証券報告書のテキスト情報を連結し、名詞を抽出した文書 d_n に対してTF-IDF法を用いることで、企業 n ごとに単語 w_k , ($k = 1, 2, 3, \dots, K$) の重要度を表す $tfidf_{w_k, d_n}$ のコーパス \mathbf{C}_n を獲得する。

tf_{w_k, d_n} = 文書 d_n 中の単語 w_k の出現頻度

$$idf_{w_k} = \log \frac{1 + N}{\text{単語 } w_k \text{ が登場する文書数}}$$

$$tfidf_{w_k,d_n} = tf_{w_k,d_n} \times idf_{w_k}$$

$$\mathbf{C}_n = ((w_1, tfidf_{w_1,d_n}), \dots, (w_K, tfidf_{w_K,d_n}))$$

次に、コーパス \mathbf{C}_n に対して、単語の分散表現と $tfidf_{w_k,d_n}$ の値で加重平均することによって、企業 n のビジネスモデルや事業内容を表すベクトル表現 \mathbf{b}_n を獲得する。

$$\mathbf{b}_n = \sum_{k=1}^K \mathbf{W2V}(w_k) \times tfidf_{w_k,d_n} \quad (4.1)$$

ただし、 $\mathbf{W2V}$ は単語から分散表現を獲得するニューラルネットワークの関数である。本研究では、簡単のため Word2Vec [109] を用いるが、FastText [110], ELMo [111] や BERT [112] といった他の単語の分散表現推定手法を用いる拡張性を持つ。

4.2.3 事業内容のベクトル表現を用いたポートフォリオの共分散行列の予測

次に、事業内容のベクトル表現 \mathbf{b}_n を用いたポートフォリオの共分散行列 $\hat{\Sigma} \in \mathbf{R}^{N \times N}$ を推定する。[39] は、標本共分散行列 \mathbf{S} と、 β を用いて市場との相関を表現した単一指標に基づく行列との加重平均によってポートフォリオの共分散行列をより精緻に推定できることを示した。本研究でも、[39] を参考に、標本共分散行列 \mathbf{S} と企業間の類似度を \mathbf{b}_n を用いた Cos 類似度で表現した行列 $\mathbf{B} \in \mathbf{R}^{N \times N}$ の要素の加重平均によってポートフォリオの共分散行列を予測する。

ポートフォリオ最適化では、標本共分散行列 \mathbf{S} は以下のように計算される。

$$\begin{aligned}
 \mathbf{S} &= \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{pmatrix} \\
 &= \begin{pmatrix} \text{Var}(R_1) & \cdots & \text{Cov}(R_1, R_N) \\ \vdots & \ddots & \vdots \\ \text{Cov}(R_N, R_1) & \cdots & \text{Var}(R_N) \end{pmatrix}
 \end{aligned} \tag{4.2}$$

ただし、 Var は分散、 Cov は共分散を表し、 R_n は企業 n の期待収益率を表す。

企業間の類似度を \mathbf{b}_n を用いた Cos 類似度で表現した行列 \mathbf{B} は、以下のように計算する。

$$\begin{aligned}
 \mathbf{B} &= \begin{pmatrix} r_{11} & \cdots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{N1} & \cdots & r_{NN} \end{pmatrix} \\
 &= \begin{pmatrix} \text{Cos}(\mathbf{b}_1, \mathbf{b}_1) & \cdots & \text{Cos}(\mathbf{b}_1, \mathbf{b}_N) \\ \vdots & \ddots & \vdots \\ \text{Cos}(\mathbf{b}_N, \mathbf{b}_1) & \cdots & \text{Cos}(\mathbf{b}_N, \mathbf{b}_N) \end{pmatrix}
 \end{aligned} \tag{4.3}$$

そして、標本共分散行列 \mathbf{S} と企業の事業内容の類似度を表す行列 \mathbf{B} との加重平均を以下のように計算し、ポートフォリオの共分散行列 $\hat{\Sigma}$ を推定する。

$$\begin{aligned}
 \hat{\Sigma} &= \mathbf{S} \otimes \mathbf{I} + \gamma \mathbf{S} \otimes (\mathbf{J} - \mathbf{I}) + \delta \mathbf{B} \otimes (\mathbf{J} - \mathbf{I}) \\
 &= \begin{pmatrix} \sigma_{11} & \cdots & \gamma \sigma_{1N} + \delta r_{1N} \\ \vdots & \ddots & \vdots \\ \gamma \sigma_{N1} + \delta r_{N1} & \cdots & \sigma_{NN} \end{pmatrix}
 \end{aligned} \tag{4.4}$$

ただし、 $\mathbf{I} \in \mathbf{R}^{N \times N}$ は単位行列、 $\mathbf{J} \in \mathbf{R}^{N \times N}$ は要素が全て1の行列、 γ, δ は正規化係数である。本研究では、標本共分散行列の対角成分の数値が大きくなりやすい性質を考慮しながら、 \mathbf{S} の絶対値の和 $|\mathbf{S}|$ と \mathbf{B} の絶対値の和 $|\mathbf{B}|$ を用いて、

$$\gamma = 0.5, \quad \delta = \frac{|\mathbf{S}|}{|\mathbf{B}|} \times 0.3$$

と計算する。

4.2.4 ポートフォリオの構築

ここでは、期待収益率と共分散行列を用いてポートフォリオを構築する手法を説明する。期待収益率の推定値を $\hat{\mu}$ とすれば、推定した $\hat{\Sigma}$ を用いて、最適なポートフォリオの割合 $\omega \in \mathbf{R}^N$ は、

$$\omega = \frac{1}{\lambda} \hat{\Sigma}^{-1} \hat{\mu} \quad (4.5)$$

とかける [113].

ただし、 $\lambda > 0$ はリスク回避度である。

4.3 実験

4.3.1 実験設定

実験対象

本研究ではまず、日本株式市場に上場している企業の株価情報や投資家向け資料の情報を用いて、将来のポートフォリオの共分散行列を予測する実験を行い、提案手法の精度を評価する。過去の株価情報だけでなく、決算短信や有価証券報告書などの投資家向け資料から得られる企業のビジネスモデルや事業内容を反映し、企業間のリスクを推定する提案手法が、将来のポートフォリオの共分散行列をより精緻に予測できることを検証する。本実験では、TOPIX 500 を構成する銘柄を

対象に，共分散行列を予測する．

その後，予測された共分散行列を用いて，ポートフォリオ最適化を行い，構築されたポートフォリオのパフォーマンスを評価する．企業のビジネスモデルや事業内容を反映した共分散行列を用いることで，より優れたポートフォリオを構築できることを検証する．

データセット

実験には，2021年5月31日付でTOPIX 500の構成銘柄であり，かつ2016年1月1日から2020年12月30日まで上場しており株価が取得可能な銘柄を対象に，それぞれの企業の株価データ，決算短信，有価証券報告書を用いる．

株価データは，2016年1月1日から2020年12月31日までの日足データを用いる．なお，途中で株式分割により発行株式数に変化があった場合は，株式分割後の発行株式数に対して前の株価を割り戻すことで，調整する．

決算短信，有価証券報告書は，2016年1月1日から2018年12月31日までに発表されているPDFデータに記述されているテキスト情報の全てを用いる．

評価方法

共分散行列の予測においては，2016年1月1日から2018年12月31日までのデータを用いて，2019年1月1日から2020年12月31日までの標本共分散行列の予測を行い，RMSEを用いて評価する． $\hat{\Sigma}$ を予測値， Σ を実際の値とすると，RMSEは以下の式で得られる．

$$RMSE = \sqrt{\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (\hat{\sigma}_{i,j} - \sigma_{i,j})^2}$$

ただし、 N は企業数、 $\hat{\sigma}_{i,j}$ は $\hat{\Sigma}$ の行列の要素、 $\sigma_{i,j}$ は Σ の行列の要素である。

ポートフォリオのパフォーマンスを評価は、構築されたポートフォリオに対して、リターン (AR)、リスクを表す標準偏差 (AVOL)、リスクあたりのリターンを表す Sharpe ratio (ASR=AR/AVOL)、ポートフォリオの全期間での最大の下落幅を表す最大ドローダウン (MDD) を用いる。

4.3.2 比較モデル

実験の目的は、過去の株価情報だけでなく、企業のビジネスモデルや事業内容を反映し、企業間のリスクを推定する提案手法が、将来の共分散行列をより精緻に予測できること、そしてその共分散行列を用いてより優れたポートフォリオを構築できることを検証することである。

提案手法の評価の正当性を保つために、提案手法をベースに、既存研究との差分にあたる部分の構造を変えた以下のモデルを比較手法とする。期待収益率 $\hat{\mu}$ も、全ての比較手法で過去の期待収益率 μ を同様に用いる。

1. SCM: 過去の標本共分散行列 (Simple Covariance Matrix) \mathbf{S} を将来の共分散行列とするモデル
2. β -SCM: 標本共分散行列 \mathbf{S} と、 β を単一指標として用いた行列を加重平均するモデル [39]
3. Sector-BCM: 式 (4.3) において、事業のベクトル表現 \mathbf{b}_n の代わりに、企業ごとに定義されている業界区分を表す one-hot ベクトルを用いたモデル

4. LSI-BCM: TF-IDF 法によって取得したコーパス \mathbf{C}_n に対して LSI (Latent Semantic Indexing, 以下 LSI と呼ぶ) [114] を用いるモデル

SCM モデル

企業のビジネスモデルや事業内容を用いず、過去の標本共分散行列 \mathbf{S} を、将来の共分散行列として予測するモデル。

$$\hat{\Sigma} = \mathbf{S} = \begin{pmatrix} \text{Var}(R_1) & \cdots & \text{Cov}(R_1, R_N) \\ \vdots & \ddots & \vdots \\ \text{Cov}(R_N, R_1) & \cdots & \text{Var}(R_N) \end{pmatrix}$$

β -SCM モデル

企業のビジネスモデルや事業内容を用いず、[39] を参考に標本共分散行列 \mathbf{S} と、 β を用いて市場との相関を表現した単一指標に基づく行列との加重平均するモデル。

企業 n の期待収益率を、CAPM [7] を用いて以下のように表す。

$$E[R_n] - r_f = \beta_{nm}(E[R_m] - r_f)$$

ただし、 $E[R_n]$ は企業 n の期待収益率、 r_f は無リスク資産の利子率、 $E[R_m]$ は市場ポートフォリオの期待収益率、 β_{nm} はマーケットリスクプレミアムに対する企業 n のリスクプレミアムの感応度であり、

$$\beta_{nm} = \text{Cov}(R_n, R_m) / \text{Var}(R_m)$$

と計算される。

この場合、 β を用いて市場との相関を表現した単一市場に基づく行列 \mathbf{S}_β は以下のように計算される。

$$\mathbf{S}_\beta = \text{Var}(R_m) \begin{pmatrix} \beta_{1m}\beta_{1m} & \cdots & \beta_{1m}\beta_{Nm} \\ \vdots & \ddots & \vdots \\ \beta_{Nm}\beta_{1m} & \cdots & \beta_{Nm}\beta_{Nm} \end{pmatrix}$$

そして、標本共分散行列 \mathbf{S} と β を単一指標として用いた共分散行列 \mathbf{S}_β との加重平均を以下のように計算し、ポートフォリオの共分散行列 $\hat{\Sigma}$ を推定する。

$$\hat{\Sigma} = \gamma\mathbf{S} + \delta\mathbf{S}_\beta$$

ただし、 γ 、 δ は正規化係数であり、本実験では

$$\gamma = \delta = 0.5$$

と計算する。

Sector-BCM モデル

式(4.3)において、事業のベクトル表現 \mathbf{b}_n の代わりに、東証33業種分類⁶に基づいて企業ごとに定義されている業種分類を表す one-hot ベクトル $\mathbf{b}'_n \in \mathbf{R}^{33}$ を用いたモデル。

$$\mathbf{B} = \begin{pmatrix} \text{Cos}(\mathbf{b}'_1, \mathbf{b}'_1) & \cdots & \text{Cos}(\mathbf{b}'_1, \mathbf{b}'_N) \\ \vdots & \ddots & \vdots \\ \text{Cos}(\mathbf{b}'_N, \mathbf{b}'_1) & \cdots & \text{Cos}(\mathbf{b}'_N, \mathbf{b}'_N) \end{pmatrix}$$

⁶<https://profile.yahoo.co.jp/>

LSI-BCM モデル

式(4.1), 式(4.4)の代わりに, TF-IDF 法によって取得したコーパス C_n に対して特異値分解を行うことで, 次元圧縮された行列式を利用する LSI [114] モデル.

対象としている企業 $n = 1, 2, \dots, N$ に対して, コーパス C_n を連結した行列 C を得る.

$$C = (C_1, C_2, \dots, C_N)$$

C に対して特異値分解を行い, 以下のように2つの直行行列と特異値からなる対角行列の内積で表す.

$$C = TS''D^T$$

そして, 以下のように B を求める.

$$B = C^T C = DS''^2 D^T$$

4.3.3 実験結果

共分散行列の予測結果

TOPIX 500 を構成する銘柄を対象に, 2019 年の共分散行列の予測の結果, 及び 2020 年の共分散行列の予測の結果は, 以下の表 4.1, 表 4.2 の通り.

表 4.1, 表 4.2 から分かるように, 検証期間において, 提案手法が他の手法よりも精緻に共分散行列を予測できている.

表 4.1: 2019年1月1日から2019年12月31日までの共分散行列の予測における提案手法のRMSE比較

	RMSE
SCM	0.0966
β -SCM	0.0726
Sector-BCM	0.1539
LSI-BCM	0.0770
Deep-BCM	0.0712

表 4.2: 2020年1月1日から2020年12月31日までの共分散行列の予測における提案手法のRMSE比較

	RMSE
SCM	0.0649
β -SCM	0.0439
Sector-BCM	0.1430
LSI-BCM	0.0504
Deep-BCM	0.0430

Sector-BCMを除いた他の手法は、単純に標本共分散行列を用いるSCMよりも精度が高く、提案手法を含む標本共分散行列と他の行列の加重平均を用いることの有用性 [39] も確認できた。

業界分類の情報を用いたSector-BCMは、ベンチマークとなるSCMの精度を上回ることができなかった。業界分類のone-hotベクトルをそのまま活用すると、同業界内での類似度や業界間での類似度を捉えることができず、むしろ過去の株価に反映されていた情報まで欠落させてしまった可能性がある。

LSI-BCMは、SCMよりも精度高く共分散行列を予測することができたが、単語の分散表現を用いてより細かく企業のリスク類似度を表現することを試みたDeep-BCMや、 β -SCMの精度を上回ることができていない。

ポートフォリオのパフォーマンス評価結果

予測された共分散行列を用いて構築されたポートフォリオの収益率の推移は、以下の図 4.2 の通り。

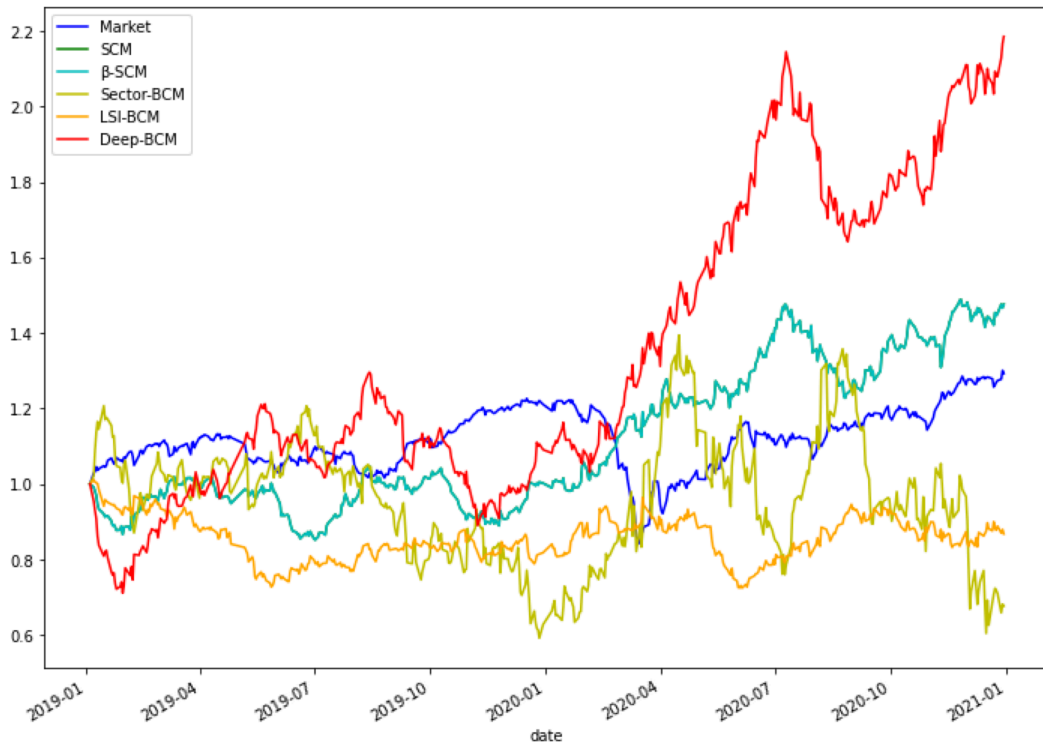


図 4.2: 各手法によって構築されたポートフォリオの収益率の推移

各手法のポートフォリオのパフォーマンスは、以下の表 4.3 の通り。

表 4.3: 各手法のポートフォリオのパフォーマンス評価

	ASR	AR	AVOL	MDD
Market	0.1601	0.0479	0.2369	0.3105
SCM	1.5645	0.4224	0.2636	0.0990
β -SCM	1.5628	0.4222	0.2638	0.0990
Sector-BCM	-0.6673	-0.3256	0.5029	0.5662
LSI-BCM	-0.4889	-0.1324	0.2912	0.2839
Deep-BCM	2.6362	1.1905	0.4478	0.3133

図4.2, 図4.3から分かるように, 検証期間において, 提案手法はSCMや β -SCMよりも優れたSharpe レシオを示し, 他手法よりも優れたポートフォリオを構築できていることが分かる.

また, ポートフォリオのパフォーマンスは, 図4.2で示される共分散行列の予測精度が反映されていることも確かめられた.

本研究の目的は, 共分散行列の予測を精緻に行うことで, より優れたポートフォリオを構築することであり, 期待収益率の予測は目的としていない. そこで, 追加の実験として, 式4.5の μ に将来の期待収益率を既知として用いて, 予測した共分散行列によって構築されたポートフォリオのパフォーマンスを表4.4のように評価した.

表 4.4: 期待収益率が既知の場合の各手法のポートフォリオのパフォーマンス評価

	ASR	AR	AVOL	MDD
SCM	8.8634	5.4975	0.6191	0.1139
β -SCM	8.8589	5.4968	0.6194	0.1137
Sector-BCM	-4.2826	-2.2345	0.5241	2.2629
LSI-BCM	3.9589	1.1248	0.2816	0.1744
Deep-BCM	9.6163	10.3374	1.0739	0.1443

表4.4から分かるように, 検証期間において, 期待収益率が既知の場合も提案手法は他手法よりも優れたポートフォリオを構築できていることが分かる.

4.4 分析

4.4.1 企業の事業内容のベクトル表現の分析

ビジネスモデルや事業内容の多様化によって, 既存の業界区分やセグメント情報では企業の事業内容を適切に表現できていない可能性がある. 提案手法を用い

て獲得されたベクトル表現 \mathbf{b}_n が、企業の事業内容をどのように表現し、業界分類を用いた one-hot ベクトルや LSI を用いたトピックモデルによるベクトル表現とどのように異なるかを分析する。

新たなビジネスモデル・事業内容に対する既存の業界区分の限界の考察

インターネットや AI といった技術革新や、COVID-19 による社会環境の変化に伴って、多様化するビジネスモデル・事業内容を既存の業界区分では表現することができない可能性がある。

技術革新が進む中で、様々な業界がデジタルを前提とした事業モデルにシフトしつつある。例えば、ラクスル(4384)は、印刷業界に対して、インターネットを用いたマッチングプラットフォームを事業の一つとして展開している。ラクスル(4384)は、東証 33 業種分類では「情報・通信」に分類されるが、印刷事業をインターネットを通じて行っている会社とも捉えることができる。図 4.3 のように、「情報・通信」や「サービス業」に分類される企業は日本株式市場 3,400 企業のうち 1,000 企業を超えている。業界既存の業界区分では「情報・通信」や「サービス業」と区分されても、実際の事業は特定の業界を対象にした事業である可能性があり、そうした事業内容は既存の業界区分では表現することができない。

東証 33 業種分類の「情報・通信」及び「サービス」と区分されるそれぞれ約 500 の企業に対して、事業内容を表すベクトル表現 \mathbf{b}_n を用いてクラスタリングを行った。

「情報・通信」に区分される企業は、図 4.4 のように Entertainment, System Integrated, Medical ICT, Internet Media の大きく 4 つのクラスタに分類された。同じ「情報・通信」でも、System Integrated や Medical ICT のような IT ソリューション企業も存在すれば、広告メディアやエンターテインメント事業をインターネットで行う企業も存在し、その性質は異なる。

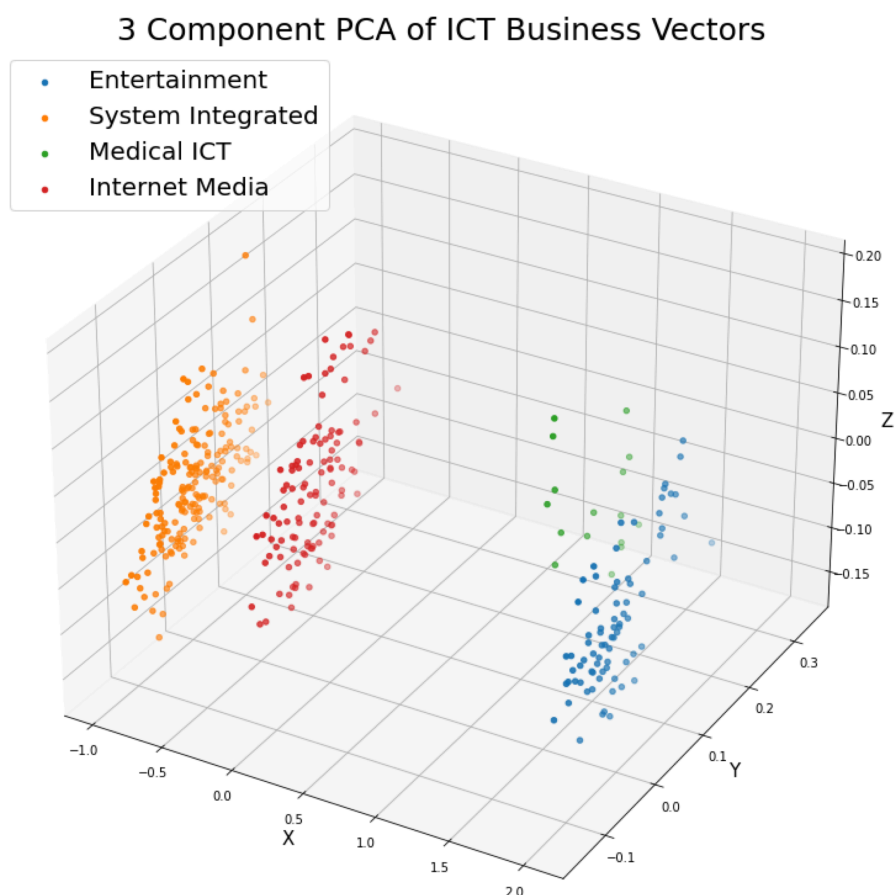


図 4.4: 「情報・通信」に属する企業の事業内容を表すベクトル表現の3次元空間へのマッピングの結果

CRI・ミドルウェア (3698) といった映像・音響関連の事業を展開する企業との類似度が高かった。ソニーの売上に関して、業界区分の「電子機器」に該当する v) エレクトロニクス・プロダクツ&ソリューションだけでなく、i) ゲーム&ネットワークサービス, ii) 音楽, iii) 映画 セグメント事業の売上も大きく、事業のベクトル表現 \mathbf{b}_n は多角化するソニーの事業の実態を表現していると考えられる。楽天グループ (4755) は、Z ホールディングス (4689), GMO インターネット (9449), ディー・エヌ・エー (2432), GMO ペイメントゲートウェイ (3769), フィスコ (3807), クレディセゾン (8253) といった企業との類似度が高かった。 \mathbf{b}_n は、EC 事業だけでなく、クレジットカードや銀行、証券と金融の事業領域へ拡大する楽天グループの事業内容を表現していると考えられる。

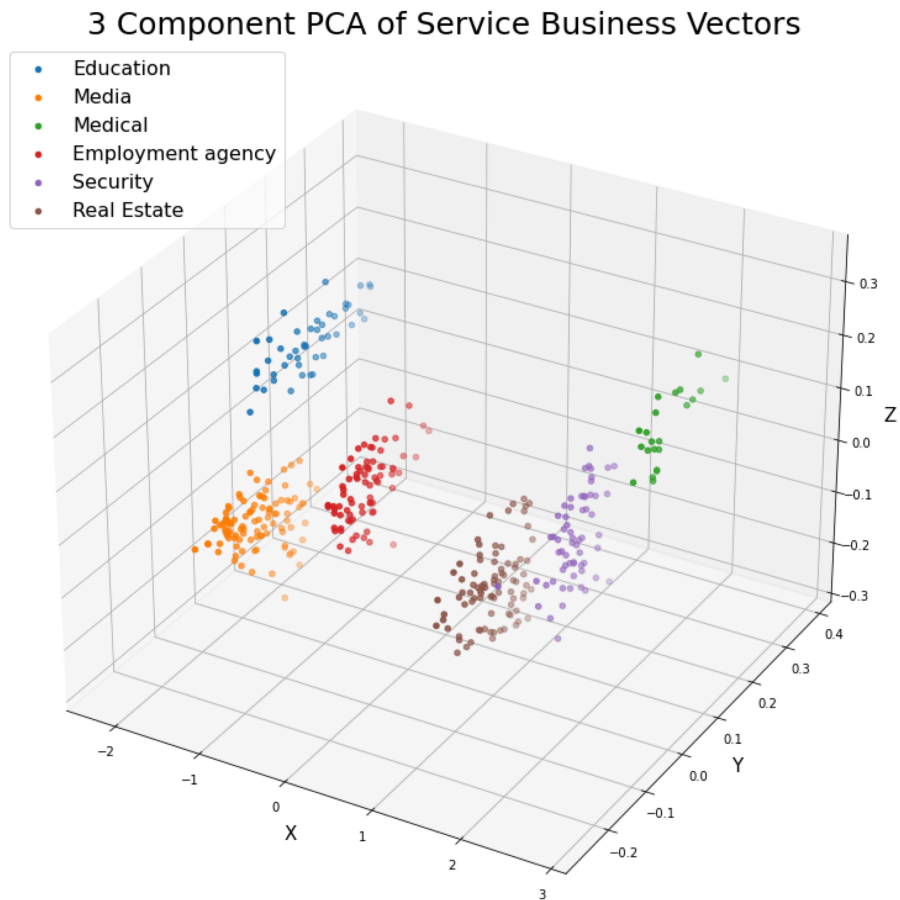


図 4.5: 「サービス業」に属する企業の事業内容を表すベクトル表現の3次元空間へのマッピングの結果

ビジネスモデルや事業内容を表すベクトル表現 \mathbf{b}_n とその導出の妥当性の検証

次に、ビジネスモデルや事業内容を表すベクトル表現 \mathbf{b}_n とその導出の妥当性について検証する。提案手法では、単語の分散表現を用いて \mathbf{b}_n を導出したが、図 4.3 で比較手法の1つとして用いた LSI-BCM の手法におけるトピック分析により、事業内容を表すベクトル表現 \mathbf{b}_n を獲得する可能性も考えられる。そこで、提案手法である Deep-BCM と、LSI-BCM の手法それぞれで導出される事業内容を表すベクトル表現 \mathbf{b}_n を比較する。

LSI-BCM と Deep-BCM の手法それぞれで導出される事業内容を表すベクトル表現 \mathbf{b}_n の分布を定量的に比較するため、クラスサイズを 10 とした Nearest Neighbor 法によるクラスタリングを行い、各クラスに含まれる企業の事業内容のベクトル表現の平均と、二乗誤差の平均を算出した。二乗誤差の平均は LSI-BCM の手法の場合 42.90、Deep-BCM の手法の場合 27.79 であり、Deep-BCM の手法の方が小さかった。一般的に、二乗誤差の平均が小さい方がクラスタリングがうまくいっていると考えることができる。Deep-BCM の手法による事業内容を表すベクトル表現 \mathbf{b}_n によって、より企業の分類がうまくいっていることが示唆される。

LSI-BCM と Deep-BCM で導出されるベクトル表現 \mathbf{b}_n を PCA 法によって 3 次元空間へ圧縮し、Nearest Neighbor 法による 10 個のクラスを可視化した結果は以下の図 4.6、図 4.7 の通り。

図 4.6、図 4.7 をみると、Deep-BCM を用いて導出された事業内容を表すベクトル表現の 3 次元空間へのマッピングの結果を表す図 4.7 の方が、それぞれのクラスが均等に分類されていると考えられる。

4.4.2 企業のビジネスモデル・事業内容と株価パフォーマンスの関係の分析

ビジネスモデルや事業内容の多様化に伴い、企業が抱える事業機会やリスクも多様化していると考えられる。既存の業界区分や事業規模に関わらず、ビジネスモデルや事業内容そのものの性質に応じて、期待収益率やリスクも異なると想定される。

近年、Software as a Service(以下、SaaS) という新しいビジネスモデルを選択する企業が存在している。従来の売り切り型のモデルと違って、SaaS はソフトウェアを利用する権利をサブスクリプション型で提供し、ユーザの利用体験価値の改善が収益に直結するという点で、ユーザフレンドリーなビジネスモデルであるとされる。投資家の視点からは、SaaS は利用者数や利用単価、解約率といった経営

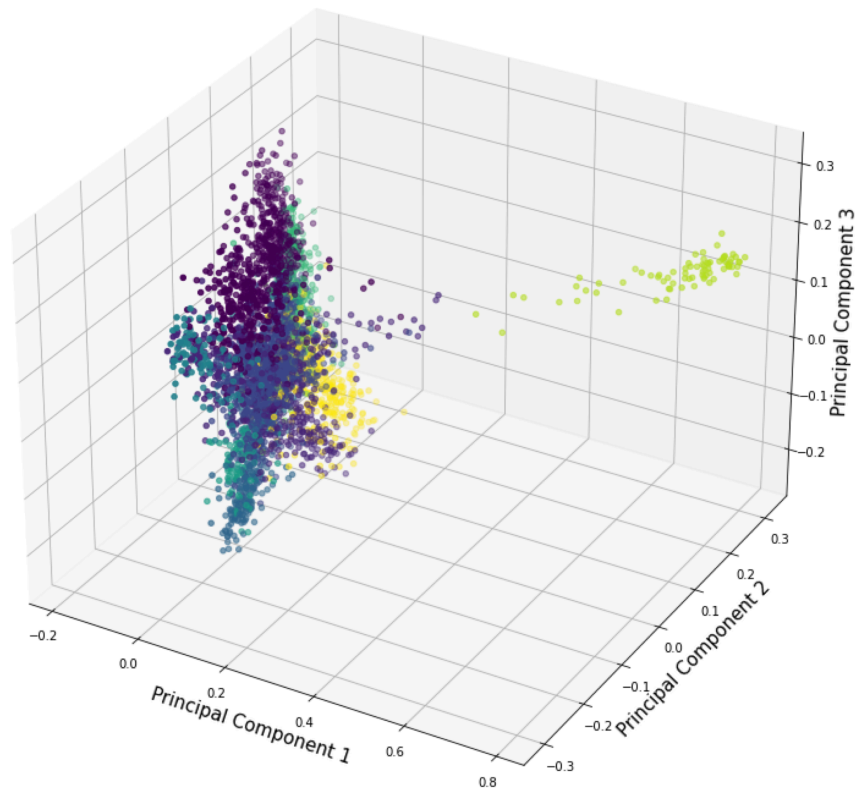


図 4.6: LSI-BCM のトピック分析によって導出された事業内容を表すベクトル表現の 3次元空間へのマッピングの結果

指標を管理することで、将来の安定的な収益を予測しやすいという点から、リスクに強いビジネスモデルとして評価されている。

グローバル化に伴い顧客チャネルやサプライチェーンを国内外に構築する企業も増えてきている。そうした企業は政治的なリスクや、COVID-19 のようなパンデミックによって顧客チャネルやサプライチェーンが分断されるリスクを抱えている。

提案手法によって導出される事業内容を表すベクトル表現 \mathbf{b}_n を用いることで、既存の業界区分や事業分類に捉われず、任意のビジネスモデル・事業内容に応じた期待収益率や分散・共分散の分析が可能になると期待される。

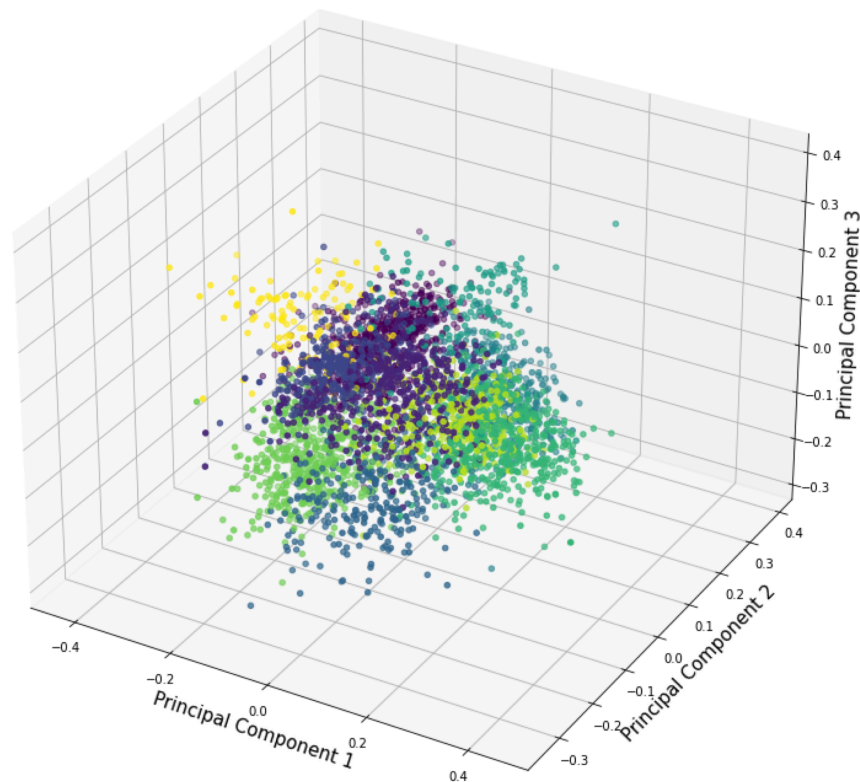


図 4.7: Deep-BCM を用いて導出された事業内容を表すベクトル表現の 3次元空間へのマッピングの結果

事業内容を表すベクトル表現 \mathbf{b}_n を用いて、日本株式市場の上場企業約 3,400 企業に対して、SaaS のビジネスモデルを持つ企業と、事業をグローバル展開している企業を抽出し、パフォーマンスの比較を行う。

SaaS 企業として、事業内容を表すベクトル \mathbf{b}_n 空間から、「SaaS」という単語の分散表現と Cos 距離が近い上位 50 社を抽出した。グローバル企業として、事業内容を表すベクトル \mathbf{b}_n 空間から、「海外」、「グローバル」、「輸出」、「輸入」という 4 つの単語の分散表現との Cos 距離が近い上位 500 社を抽出した。

SaaS 企業、グローバル企業、上場企業全体の平均の 2019 年 1 月 1 日から 2020

年12月31日までの株価のパフォーマンスは以下の図4.8の通り。



図 4.8: ビジネスモデル・事業内容ごとの株価パフォーマンスの比較. SaaS 企業は相対的にパフォーマンスが高く, COVID-19 の感染拡大後も急速な回復をしている. 顧客チャネルやサプライチェーンを国内外に構築するグローバル企業は, COVID-19 後のパフォーマンスが相対的に低迷している.

特に, COVID-19 が流行する以前の 2020 年 2 月までと, 流行が始まった 2020 年 3 月からのパフォーマンスは, SaaS 企業とグローバル企業では開きがある. SaaS 企業は, 顧客が国内におりかつインターネット上で営業活動が完結するプロダクトを持っている企業も多く, またリカーリングリベニューがあるため, COVID-19 による業績へのネガティブな影響は少なく, 株価が回復するのも相対的に早かったと考えられる. 一方グローバル企業は, COVID-19 の流行でサプライチェーンが分断され, 海外の工場では感染者の発生に伴い工場が停止するなどして, 業績への営業も大きかったため, 株価の落ち込みが大きく, また株価の回復も遅れていると考えられる.

本研究で獲得された事業内容を表すベクトル表現 \mathbf{b}_n を用いることで, ビジネス

モデルや事業内容の性質に応じた期待収益率やリスクの分析可能性が広がった。

4.5 終わりに

インターネットやAIといった技術革新や、COVID-19による社会環境の変化に伴って、企業のビジネスモデルや事業内容が多様化している。そこで本研究では、株価情報だけを考慮するのではなく、ニューラルネットワークを用いて企業のビジネスモデルや事業内容のベクトル表現を獲得し、ポートフォリオの共分散行列を予測する手法を提案した。

実験を通して、企業のビジネスモデルや事業内容のベクトル表現 \mathbf{b}_n を用いる提案手法が、将来の共分散行列をより精緻に予測できることを示した。また、提案手法を用いて予測される共分散行列を用いて構築されたポートフォリオは、ビジネスモデルや事業内容を考慮しない場合のポートフォリオよりも優れたパフォーマンスを出すことを示した。さらに、提案手法を用いて導出されるベクトル表現 \mathbf{b}_n を用いることで、既存の業界区分やセグメント情報では適切に表現しきれない企業情報を抽出できることや、ビジネスモデルや事業内容の性質に応じた期待収益率やリスクの分析可能性が高まっていることを示唆した。

本研究では Word2Vec [109] を用いて単語の分散表現を導出したが、Deep Learning 分野の研究の進展に伴い、さらなる手法の改善も期待される。さらに、本研究の目的を超えて、企業のビジネスモデルや事業内容のベクトル表現を用いて、期待収益率の予測のモデリングや、企業価値評価を行う研究へ進展していくことを期待している。

第5章 ソーシャルメディアにおける影響関係から金融市場に対する作用のモデル化と分析

本章では，投資家心理に着目し，ヒトの認知処理能力では取り扱うことが難しかったソーシャルメディアの影響関係から金融市場への作用を，深層学習を用いてモデル化することで，価格変動をより精緻にかつ安定して予測可能な手法を提案する。

5.1 はじめに

金融市場の観測や予測をする際，ソーシャルメディアを実世界の射影と捉え，観測・予測に有用な情報源として用いる手法が提案されている [53–56]。ソーシャルメディア上の情報は，実世界で起こっていること，それに対する人々の反応を示す情報であり，全体として社会を観測するセンサと考えることができる [9]。こうした機能は，ソーシャルセンサ [63] と呼ばれ，実世界における社会での事象やトレンドの検知を可能にし，我々の実世界の観測と予測の可能性を広げた。投資家が何を認知し，どのような感情を抱いているかを反映するセンサとしてソーシャルメディアを用いることで，金融市場の観測や予測をより精緻に行うことが可能になっている [53]。

一方で，近年，ソーシャルメディア内での相互作用が，投資家の心理状態の変化を引き起こし，金融市場の価格変動の決定要因の一つになっている可能性がある [74]。ソーシャルメディアにおいて他者の投稿を閲覧し影響を受けることで，特

定の金融資産に対する認知や心理状態を変化させ、投資行動を行っている投資家が存在する。[74]では、ある特定のツイッターユーザのツイートによって株価が変動している可能性が言及されている。本研究は、ソーシャルメディアが、金融市場の価格変動に働きかけるアクチュエータとしても機能しているという仮説に基づく。

暗号資産のような需要と供給がより投資家心理に左右されやすい金融資産の登場に伴い、ソーシャルメディア内の相互作用と心理状態の変化が金融市場に作用する事象をモデル化し、分析する重要性は高まっている。暗号資産には、ビットコイン、イーサ(イーサリアムの内部通貨、以下イーサリアムと呼ぶ)、リップルなど様々な暗号資産があり、ビットコインは2017年初頭では\$900ほどであったのに対して、2017年12月には\$20,000近くまで価格が上昇し、注目を集めた。暗号資産はマクロ経済に基づくファンダメンタルズ要素がないとされ[56]、将来のキャッシュフローモデル[58]、購買力平価[59, 60]、および金利パリティ[61, 62]のような既存の経済理論や金融理論によって、暗号資産の価格変動を説明することは困難である。暗号資産の需要と供給は、相対的に投資家心理により左右されると推察される。

本研究は、ソーシャルメディアにおける影響関係から金融市場に対する作用をモデル化し、暗号資産をはじめとする金融資産の価格変動をより精緻にかつ安定して予測することを目的とする。そして、モデルを用いた解析によって、ソーシャルメディアにおける影響関係から金融市場に対する作用の分析を行う。

ソーシャルメディアにおけるユーザの影響関係とその金融市場に対する作用は、直接観測することができず、時間を通して相互に複雑に関連しているため、モデル化し分析することは困難であった。金融市場に対して作用をもたらすソーシャルメディア内の影響関係は、投稿やいいね、リツイートといった観測可能な情報拡散を取り扱う information diffusion [78]の研究と異なる。他のユーザの投稿の閲覧に伴う影響関係を含めた、直接観測できないユーザの認知や心理状態の変化を

考慮する必要がある。また、他のユーザから繰り返し影響を受けたユーザが、時間を経て他者に影響を与えるという可能性もあり、ソーシャルメディアにおける相互作用とユーザの状態、及び金融市場への影響は、時間を通して複雑に関連している。

そこで本研究は、ソーシャルメディアにおける影響関係から金融市場に対する作用を定式化し、end-to-endのニューラルネットワークであるRecurrent Neural Network(以下RNNと呼ぶ)を用いてモデル化する手法(図5.1)を提案する。金融市場で観測される値を予測対象とした損失関数を持つend-to-endのRNNモデルを学習させることで、直接観測できず、実際の金融資産の価格変動に寄与する相互作用や、時間を通して複雑に変化するユーザの認知や心理状態を抽出・分析ができるようになることが期待される。

本稿では、金融市場の予測のタスクに対して、提案手法のモデルと、ソーシャルメディアにおける影響関係から金融市場に対する作用を反映していない複数のモデルを比較検証した。検証実験では、サンプリングされた投資家のツイッターアカウント9,900人の1,133,801個の有向エッジを持つグラフネットワーク及び2017年1月1日から2017年12月31日までの9,842,699件のツイートを集積し、暗号資産市場と日本の株市場の2つの金融市場の価格データを用いた。その後、提案手法のモデルのパラメタを解析することによって、ソーシャルメディアにおける影響関係から金融市場への作用に関する分析を行った。

本研究の貢献は、以下の通りである。

- ソーシャルメディアの影響関係から金融市場への作用をモデル化する手法を提案し、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻にかつ安定して予測可能であることを示したこと。
- 提案手法のモデルによって、ソーシャルメディアの影響関係から金融市場への作用の分析可能性を高めたこと。その分析の中で、ソーシャルメディア内のネットワークが直接接続されていなくとも、時間を通して影響を与える関

係になることや、金融資産の性質によってソーシャルメディアから金融市場への影響の強さが異なることを示唆したこと。

5.2 提案手法

5.2.1 ソーシャルメディアの影響関係から金融市場への作用の定式化

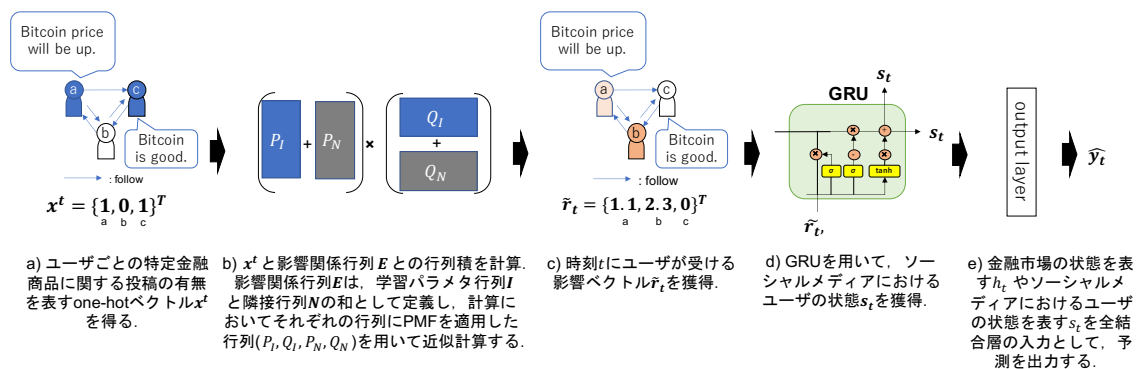


図 5.1: end-to-end のネットワークを用いたモデルの概要

ソーシャルメディアの一部のユーザは、他者の投稿を閲覧し、特定の金融資産に対する認知や心理状態を変化させ、意思決定を行っている。そうして新たな投資行動を引き起こし、結果として金融市場が変動していると推察できる。本研究では、モデル化するソーシャルメディアの影響関係から金融市場への作用を以下のステップに定式化する。

1. 観測と投稿 ユーザ u_j が金融市場に関わる事象 X を観測し、事象 X に関連する投稿をソーシャルメディアに行う。
2. 影響 ユーザ u_j の投稿を閲覧したユーザ u_i は、事象 X に対する認知や心理状態に影響を受け、場合によっては事象 X に関連する発信を行い、ソーシャルメディアにおいて情報拡散が起こる。
3. 認知と心理変化 ユーザ u_i や、ユーザ u_i の投稿を閲覧した他のユーザが、さらに事象 X に対する認知を高めたり、事象 X に対する心理状態を変化させる。

4. 行動 ソーシャルメディアから認知や心理状態に対して影響を受けたユーザが、実際に事象 X に関連する投資行動を行い、金融市場が変動する。

5.2.2 end-to-end のネットワークを用いたモデル化

次に、項で定式化したソーシャルメディアの影響関係から金融市場に対する作用を、end-to-end のニューラルネットワークである RNN を用いてモデル化する手法(図 5.1)を説明する。

金融市場の観測値を予測対象とした損失関数を持つ end-to-end の RNN モデルを学習させることで、実際の金融資産の価格変動に寄与する相互作用や、時間を通して複雑に変化するユーザの認知や心理状態を抽出・分析できると期待される。

モデルの概要は図 5.1 の通り。まず、ソーシャルメディアから特定の金融資産に関連するユーザの投稿の有無を表す one-hot ベクトルを入力とし、ユーザの影響関係を表す行列 E との行列積を計算することで、時刻 t にユーザが受ける影響ベクトル r_t を獲得する。その後 RNN の一種である GRU を用いて時系列におけるユーザの認知・心理状態を表す s_t を獲得し、最終層を経て金融市場の予測を出力する。

なお、本研究では、ソーシャルメディアにおける影響関係から金融市場に対する作用をモデル化することで、金融資産の価格変動をより精緻にかつ安定して予測することを目的としている。そのため、モデルの詳細は次節以降で説明するが、ニューラルネットワークの構造や具体的な学習手法は、その性能や分析可能性を考慮し、異なるものを適用しても良い。

5.2.3 影響関係行列 E による影響関係の表現と抽出

ソーシャルメディアを用いて金融市場の予測を行う場合、リプライやリツイートなどの直接観測されるユーザだけでなく、閲覧によって引き起こされる直接観

測できないユーザの影響関係を考慮することが望ましい。

LTM [77] では、ユーザ u_j からユーザ u_i に対する影響度 $e_{j,i}$ は直接観測できないので、影響度 $e_{j,i}$ をフォロワー数などのネットワーク指標で代用するが、本研究のように時間を通したユーザの影響関係を取り扱う際に、ネットワーク指標のみを影響関係指標とすることには、以下の2つの課題がある。

1. ユーザ u_i がユーザ u_j をフォローし投稿を閲覧する場合でも、ユーザ u_i がユーザ u_j を信用していない、またはユーザ u_j が金融市場に関連のない投稿をしている場合は、金融市場に作用をもたらすほどの影響関係は実際には存在せず、相互作用の関係を過大評価する可能性がある。
2. ユーザ u_i がユーザ u_j をフォローしておらず、直接投稿を閲覧しない場合でも、ユーザ u_i の興味関心がユーザ u_j の投稿する情報と類似している場合、他のユーザを介した影響の連鎖によって、時間を通して影響関係にある可能性がある。

そこで本研究では、閲覧などによって引き起こされるユーザ u_i がユーザ u_j から受ける影響 $s_{j,i}$ を要素にもつ影響関係行列 E を end-to-end ネットワークの学習パラメタとして定義し、金融市場で観測される値を予測対象とした損失関数を用いた学習過程で表現獲得することを試みる。

また、ユーザ u_i がユーザ u_j をどのように評価し、影響を受けるかという $e_{j,i}$ を推定するタスクと、それに関連する上述した課題は、未知なる2者間の評価関係を推定するという枠組みにおいて、レコメンドアルゴリズムにおけるユーザ u_i のアイテム i_j に対する評価を推定するタスクおよびその課題に類推できる。[115]の研究では、ユーザ u_i がアイテム i_j を閲覧することと、ユーザ u_i がアイテム i_j に対して実際に影響を受けることを別々の行列で表すことで、上述した二つの課題に類似する問題の解決を試みている。そこで本研究も、対象とするソーシャルメディアのユーザ M 人のうち、ユーザ u_i がユーザ u_j の投稿を閲覧することを、ソーシャルメディアのフォロー関係を表す隣接行列 N によって表現し、ユーザ u_i がユーザ

u_j の投稿を閲覧した時に受ける影響度を、学習パラメタ行列 \mathbf{I} によって表す。そして、隣接行列 \mathbf{N} と学習パラメタ行列 \mathbf{I} の和によって、影響関係行列 $\mathbf{E} \in \mathbf{R}^{M \times M}$ を定義する。

$$\mathbf{E} = \mathbf{N} + \mathbf{I} \tag{5.1}$$

ただし、行列 \mathbf{N} の要素 $n_{i,j}$ はユーザ u_i がユーザ u_j をフォローしているか否かを表すバイナリ値であり、学習パラメタ行列 \mathbf{I} の要素 $i_{i,j}$ は閲覧などに伴いユーザ u_i がユーザ u_j から受ける影響を表現する値である。

そして、ある時刻 t において、 K 人のユーザが、対象としている金融資産に関連する投稿をしたかどうかを表した one-hot ベクトル $\mathbf{x}_t \in \mathbf{R}^M$ と影響関係行列 \mathbf{E} の行列積によって、ある時刻 t にユーザが受ける影響ベクトル $\mathbf{r}_t \in \mathbf{R}^M$ を獲得する。

$$\mathbf{r}_t = \mathbf{E}\mathbf{x}_t \tag{5.2}$$

ソーシャルメディアのような高次元のグラフを取り扱う時、対象とする K 人のユーザのすべての影響関係を学習することは計算量が膨大である上、ユーザの投稿の有無はスパースなデータであり、学習過程において過学習をする恐れもある。高次元なグラフネットワークを低次元で表現する手法として、Probabilistic Matrix Factorization [116](以下 PMF) を用いる。影響関係行列 \mathbf{E} を構成する隣接行列 \mathbf{N} に対して PMF を行い、行列 $\mathbf{P}_N \in \mathbf{R}^{M \times K}$ と行列 $\mathbf{Q}_N \in \mathbf{R}^{K \times M}$ に分解し、 \mathbf{I} に基づく学習パラメタ行列 $\mathbf{P}_I \in \mathbf{R}^{M \times K}$ と学習パラメタ行列 $\mathbf{Q}_I \in \mathbf{R}^{K \times M}$ を定義することで、式 (5.1) の影響関係行列 \mathbf{E} を以下のように近似する。

$$\mathbf{E} \sim (\mathbf{P}_I + \mathbf{P}_N)(\mathbf{Q}_I + \mathbf{Q}_N)$$

よって、式 (5.2) は、以下のように近似され表される。

$$\mathbf{r}_t \simeq (\mathbf{P}_I + \mathbf{P}_N)(\mathbf{Q}_I + \mathbf{Q}_N)\mathbf{x}_t \quad (5.3)$$

5.2.4 認知・心理の時系列内部状態の獲得

ソーシャルメディアのユーザは、他のユーザの影響を受けてからすぐに関連する投稿をしたり、投資行動を起こすとは限らない。例えば、ユーザ u_i は他のユーザ u_j の金融資産 X に対する投稿を繰り返し閲覧することによって、その金融資産 X に関して認知を深め、興味関心や心理状態が変化することで、投資を行う場合がある。この時、時間を通じた影響関係やソーシャルメディアのユーザの認知や心理状態を考慮する必要がある。

ネットワークの状態は、直接観測することができないため、HMMによって表現する手法 [76] を用いることが考えられる。しかし、ユーザの影響関係と、認知・心理状態の時間変化は相互に密接に関連すると考えられ、本研究では end-to-end のアルゴリズムであり、時系列モデルを取り扱う RNN の一種である GRU [18] を用いる。

まず、式 (5.3) で獲得された影響ベクトル \mathbf{r}_t に対して、次の処理を行い、埋め込みベクトル $\tilde{\mathbf{r}}_t \in \mathbf{R}^{hidden}$ を得る。

$$\tilde{\mathbf{r}}_t = dropout(\tanh(\mathbf{W}_{in} \frac{\mathbf{r}_t}{\sqrt{M}} + \mathbf{b}_{in}))$$

ただし、*dropout* は学習時に用いる Dropout [117] 層、 M は対象とするソーシャルメディアのユーザ数、 \mathbf{W}_{in} および \mathbf{b}_{in} は学習パラメタである。

そして、時刻 t において計算できるソーシャルメディアのユーザの認知・心理状

態を表す \mathbf{s}_t を, GRU を用いて獲得する.

$$\mathbf{s}_t = GRU(\tilde{\mathbf{r}}_t) \quad (5.4)$$

一方で, 獲得された \mathbf{s}_t と, タイムステップ t において計算できる金融市場の価格変動を表す特徴量 \mathbf{m}_t に対し, GRU を適用し, 金融市場の状態を表す特徴量 \mathbf{h}_t を獲得する.

$$\mathbf{h}_t = GRU(\tilde{\mathbf{m}}_t) \quad (5.5)$$

金融資産の特性に応じて, ソーシャルメディアの重要性が異なる可能性があり, ソーシャルメディアが金融市場の価格変動の予測にどれだけ寄与しているのかという分析の可能性を高めるため, 本研究では \mathbf{s}_t および \mathbf{h}_t を 1次元の値に変換し, Batch Normalization [118] を用いて正規化したのち最終層に入力して予測を行う.

$$s_t = \tanh(\mathbf{W}_{s-out}\mathbf{s}_t + b_{s-out})$$

$$h_t = \tanh(\mathbf{W}_{h-out}\mathbf{h}_t + b_{h-out})$$

$$\hat{y}_t = \tanh(\mathbf{W}_{out}\text{concat}(s_t, h_t) + b_{out}) \quad (5.6)$$

ただし, $\text{concat}(s_t, h_t)$ は s_t と h_t の連結ベクトルであり, \mathbf{W}_{s-out} , b_{s-out} , \mathbf{W}_{h-out} , b_{h-out} , \mathbf{W}_{out} , b_{out} は学習パラメタ, \hat{y}_t は時刻 t に出力される予測値である.

5.3 実験

5.3.1 実験設定

実験対象

本研究ではまず、ツイッターのデータを用いて、暗号資産の将来の価格変動のボラティリティを予測する実験を行い、提案手法の学習モデルの精度を評価する。ソーシャルメディアの影響関係から金融市場に対する作用をモデル化する提案手法が、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻に予測できることを検証する。本実験では、ビットコイン、イーサリアム、リップルの暗号資産市場で時価総額が大きい3つの暗号資産について、将来のボラティリティの予測を行う。予測するボラティリティは、予測する時点から特定タイムステップ経過後までの期間の、1タイムステップ前からの価格変化率の標準偏差とする。

データセット

実験には、ビットコイン、イーサリアム、リップルの暗号資産市場で時価総額が大きい3つの暗号資産の価格データ、Twitter REST API¹から取得した投資家のツイートおよびフォロワー/フォロワー関係のデータを用いる。

暗号資産市場のデータは、2017年1月1日から2017年12月31日までの1年間の価格の日足データを用いる。なお、本実験では1タイムステップの $\Delta t = 1$ は1日とする。取得したデータから、以下の要素を計算し、時刻 t における暗号資産の価格変動を表す特徴量 m_t として利用する。

1. $ChangePrice_{\Delta t}$: 時刻 t から $\Delta t = 1, 3, 5, 10, 25$ 日前からの終値変化率
2. $DeviationRate_{\Delta t}$: 時刻 t における $\Delta t = 3, 5, 10, 25$ 日間の終値移動平均からの乖離率
3. $Volatility_{\Delta t}$: 時刻 t における $\Delta t = 3, 5, 10, 25$ 日間のボラティリティ(過去 Δt 日間の、1日前からの終値変化率 $ChangePrice_1$ の標準偏差)

¹<https://dev.twitter.com/twitterkit/android/access-rest-api>

また、時刻 t から 1,3,5,10 日後までの期間の将来のボラティリティを計算し、予測対象とする。

ツイッターのデータは、[74] で言及されているある投資家ツイッターアカウントをフォローしている約 200,000 人 (2018 年 1 月時点) のユーザのうち、2017 年 1 月から 4 月までの 3 ヶ月間で 5 回以上ツイートがあるアクティブユーザで、かつフォロワーが 100 人以上いる 9,900 人のユーザのアカウントを対象とする。そして、対象アカウント 9,900 人の全てのツイート、フォローしているユーザ、フォロワーのユーザを独自に獲得した。全部で、2017 年 1 月 1 日から 2017 年 12 月 31 日までの 9,842,699 件のツイートおよび、1,133,801 個の有向エッジを持つ 9,900 人の対象ユーザのグラフネットワークを分析対象とした。

評価方法

評価方法として、暗号資産市場の将来のボラティリティの予測においては、RMSE を用いる。 \hat{y} を予測値、 y を実際の値とすると、RMSE は以下の式で得られる。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}.$$

暗号資産市場のデータは、1 月 1 日から 10 月 31 日までの約 300 レコードを学習データとし、11 月 1 日から 12 月 31 日までの 60 レコードをテストデータとして用いて評価を行う。

ハイパーパラメタ

モデルの学習には、Adam を用いて、学習率を 0.001 とする。トレーニングステップは最大 1000 とし、100 ステップごとにバリデーションスコアを確認し、バリ

デーションスコアが前回の値を下回った場合に学習を終了させる。バリデーションのためのデータセットは、学習データの時系列の最後から32ステップとした。ドロップアウト率を0.3、影響ベクトル \mathbf{r}_t の次元を 2^6 、隠れ層ノードの数を 2^4 とする。提案モデルのRNNにおいて、本検証実験では時系列長を5タイムステップとした。

5.3.2 比較モデル

表 5.1: 提案手法と比較手法のデータ利用・構造の違い

	Market	Social Sensor	LTM	Proposed Model
金融市場データ	○	○	○	○
ソーシャルメディアの投稿に関する統計量	×	○	○	○
ネットワークにおける閲覧を伴う影響関係	×	×	○	○
時間を通じた影響関係と心理状態	×	×	×	○

実験の目的は、ソーシャルメディアの影響関係から金融市場に対する作用を、end-to-endのニューラルネットワークであるRNNを用いてモデル化する手法が、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻に予測できることを検証することである。

提案手法の評価の正当性を保つために、比較手法は、提案手法のモデルをベースに、既存研究との差分の構造を変えた以下の3つのモデルとする。

1. Market: 提案手法のモデルにおいて、ソーシャルメディアから得られる情報を用いず、金融市場のデータのみを扱う時系列モデル
2. Social Sensor: 提案手法のモデルにおいて、式(5.4)にソーシャルメディアにおける投稿に関する統計量を用いた、ソーシャルセンサの時系列モデル

3. LTM: 提案手法のモデルにおいて, 式 (5.4) にソーシャルメディアの影響関係として LTM を利用した時系列モデル

Market: ソーシャルメディアから得られる情報を用いず, 金融市場のデータのみを扱う時系列予測モデル

ソーシャルメディアから取得可能なデータを用いずに, 上述した時刻 t における暗号資産の価格変動を表す特徴量 \mathbf{m}_t のみを入力として, GRU を用いた予測を行う。出力として, 以下を得る。

$$\tilde{\mathbf{h}}_t = GRU(\mathbf{m}_t)$$

$$\hat{y} = \mathbf{W}_{out}\tilde{\mathbf{h}}_t + b_{out},$$

ただし, $\mathbf{W}_{out}, b_{out}$ は学習パラメタである。

Social Sensor: ソーシャルメディアにおける投稿に関する統計量を用いた, ソーシャルセンサの時系列モデル

次に, 金融市場のデータ \mathbf{m}_t に加えて, ソーシャルメディアにおいて対象とする M 人のユーザのうち, 時刻 $t-1$ から t までの間に特定の金融資産についての言及を含む投稿したユーザの割合 num_t を用いて, 金融市場の予測を行う。

$$\tilde{\mathbf{h}}_t = GRU(\text{concat}(\mathbf{m}_t, num_t))$$

$$\hat{y} = \mathbf{W}_{out}\tilde{\mathbf{h}}_t + b_{out},$$

ただし, $\text{concat}(\mathbf{m}_t, num_t)$ は num_t と \mathbf{m}_t を連結したベクトルであり, $\mathbf{W}_{out}, b_{out}$ は学習パラメタである。

暗号資産市場に関連する投稿は, ビットコイン, イーサリアム, リップルのそれ

それぞれの暗号資産において以下のワードが含まれている投稿を対象とした。

ビットコイン: ビットコイン, Bitcoin, BTC

イーサリアム: イーサリアム, Ethereum, ETH

リップル: リップル, Ripple, XPR

LTM: ソーシャルメディアにおける閲覧による影響関係を考慮した, LTM を利用した時系列モデル

ソーシャルメディアにおける閲覧に伴う影響関係を考慮するモデルとして LTM を用いることが考えられる。LTM では, ユーザ u_j からユーザ u_i に対する影響度 $s_{j,i}$ を用いるが, 影響度 $s_{j,i}$ を直接観測することができないので, 影響度をネットワーク指標などで代用して表す必要がある。[78] において影響度をフォロワー数で表現するモデルが示されており, また information cascades のようなソーシャルメディアにおけるネットワーク構造の将来予測に関しても, ソーシャルネットワークの構造を表すネットワーク指標を入力として用いることの有用性が示されている。

s

そこで, Social Sensor モデルのように, 単にユーザの投稿割合を入力にするのではなく, その投稿を行ったユーザのネットワーク指標によって重み付けした値 num_t を入力として用いる。

$$num_t = \frac{1}{\sqrt{M}} \sum_{i=1}^n degree(user_i)$$

ただし, M は検証実験で対象としているユーザ数, n は時刻 t において特定の金融資産に関する投稿を行ったユーザの数, そして $degree(user_i)$ は $user_i$ のソーシャルメディアにおける対象としているユーザのネットワーク指標 (本研究では, フォロワー数とする) を表す。

出力として, 以下を得る。

$$\tilde{\mathbf{h}}_t = GRU(\text{concat}(\mathbf{m}_t, num_t))$$

$$\hat{y} = \mathbf{W}_{out} \tilde{\mathbf{h}}_t + b_{out},$$

ただし、 $\text{concat}(\mathbf{m}_t, \text{num}_t)$ は num_t と \mathbf{m}_t を連結したベクトルであり、 $\mathbf{W}_{out}, b_{out}$ は学習パラメタである。

5.3.3 実験結果

ビットコイン、イーサリアム、リップルの将来の価格変動ボラティリティ予測の結果は、それぞれ表 5.2, 表 5.3, 表 5.4 の通り。

表 5.2: ビットコインの将来のボラティリティ予測における提案手法の RMSE 比較

	1 日後	3 日後	5 日後	10 日後
Market	0.0230	0.0309	0.0323	0.0420
Social Sensor	0.0394	0.0416	0.0307	0.0459
LTM	0.0317	0.0323	0.0392	0.0431
Proposed model	0.0215	0.0220	0.0267	0.0414

表 5.3: イーサリアムの将来のボラティリティ予測における提案手法の RMSE 比較

	1 日後	3 日後	5 日後	10 日後
Market	0.0146	0.0200	0.0245	0.0269
Social Sensor	0.0176	0.0218	0.0244	0.0239
LTM	0.0160	0.0196	0.0253	0.0243
Proposed model	0.0184	0.0184	0.0226	0.0237

表 5.4: リップルの将来のボラティリティ予測における提案手法の RMSE 比較

	1 日後	3 日後	5 日後	10 日後
Market	0.0654	0.0813	0.0931	0.1042
Social Sensor	0.0572	0.0730	0.0957	0.0993
LTM	0.0617	0.0728	0.0900	0.0994
Proposed model	0.0627	0.0655	0.0790	0.0938

表 5.2, 表 5.3, 表 5.4 からわかるように、ほとんどの比較検証実験の場合において、提案手法が他の手法よりも精緻に将来のボラティリティの予測を行えている。ソーシャルメディアにおける影響関係から金融市場への作用をモデル化する提案

手法が、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻に予測できることを示唆する。

一方で、表5.3, 表5.4において、予測時点から1日後のイーサリアムおよびリップルのボラティリティの予測では、提案手法はMarketやSocial Sensorの手法を上回ることができなかった。予測時点直後のボラティリティに関しては、ソーシャルメディアにおけるユーザの状態が反映されるよりも、現在のボラティリティの傾向が継続しやすく、Marketのような比較的単純な手法の方が、より精緻に予測ができていたのだと推察できる。

また、Social SensorやLTMの手法は、ソーシャルメディアの情報を用いないMarketに対して、安定的に優れた予測精度を出すことができていない。3種類の暗号資産の、1, 3, 5, 10日後の予測実験の計12パターンの実験において、Social Sensorの手法は、5パターンがMarketの精度を上回ることができず、LTMの手法は、6パターンがMarketの精度を上回ることができなかった。特にビットコインの将来ボラティリティ予測では、Social Sensorの手法は、5日後の予測のみMarketの精度を上回り、LTMの手法においては、Marketの精度を上回ることができていない。

Social SensorやLTMの手法では、ソーシャルメディアの投稿の閲覧に伴う影響関係や、時間を通して変化・蓄積されるユーザの認知や心理状態を抽出・表現することができておらず、将来の予測という枠組みでは、将来に対する影響が少なく、ただ価格変動に対して反応したユーザの投稿が、かえってノイズとなってしまっている可能性がある。一方で、提案手法は、複雑でかつ直接観測されないソーシャルメディアの投稿の閲覧に伴う影響関係や、時間を通して変化・蓄積されるユーザの認知や心理状態を抽出・表現するため、そうしたノイズに対して頑強で、表5.2, 表5.3, 表5.4で示されるように、より安定して高い精度を出すことができていると考えられる。

5.4 分析

5.4.1 ソーシャルメディアの社会ネットワークにおける影響関係の抽出と分析

ソーシャルメディアにおいて、リツイートなどの情報拡散だけでなく、閲覧を伴う直接観測されない情報拡散も存在するはずである。また、フォロー/フォロワーという直接的な関係だからと言って、どれだけ影響を受けるかどうかはわからないし、時間を通してユーザの認知や心理状態が変化することで、フォロー/フォロワーという直接的な関係がなくても、影響関係にある可能性がある。

そこで、提案手法の学習モデルのユーザ間の影響度を表すパラメタ行列 E を分析することで、ソーシャルメディアにおける影響関係から金融市場への作用を考慮した、ソーシャルメディア内のユーザ同士の影響関係を分析する。

図5.2は、ビットコインの3日後のボラティリティ予測の検証実験で獲得された影響関係を表す行列 E において、サンプリングされた10アカウントの影響関係に対応するパラメタをヒートマップで表したものであり、値が大きいほど影響度が高い。行は影響を与えるユーザを表しており、列は影響を受けるユーザを表している。また、青四角で囲まれたユーザ関係は、ツイッターにおいて列のユーザが行のユーザをフォローしていることを表している。

提案手法を用いて、複雑でかつ直接観測されないソーシャルメディアの投稿の閲覧に伴う影響関係を陽に表現する手法であり、図5.2のように、誰が誰にどれだけ影響を与えているのかを獲得し、可視化することが可能である。

図5.2を見ると、青四角で囲まれておらず、フォローしていないが影響度が高く抽出されている関係があることがわかる。例えば、あるユーザ u_{2772} がユーザ u_{874} から与えられる影響度は、 $E_{(874,2772)} = 0.59$ であり、フォロー関係になかったにもかかわらず強い影響関係が見て取れる。ユーザ u_{2772} は、実際に、暗号資産市場に関する多くのつぶやきを投稿していたことが確認され、ユーザ u_{2772} はビットコイ

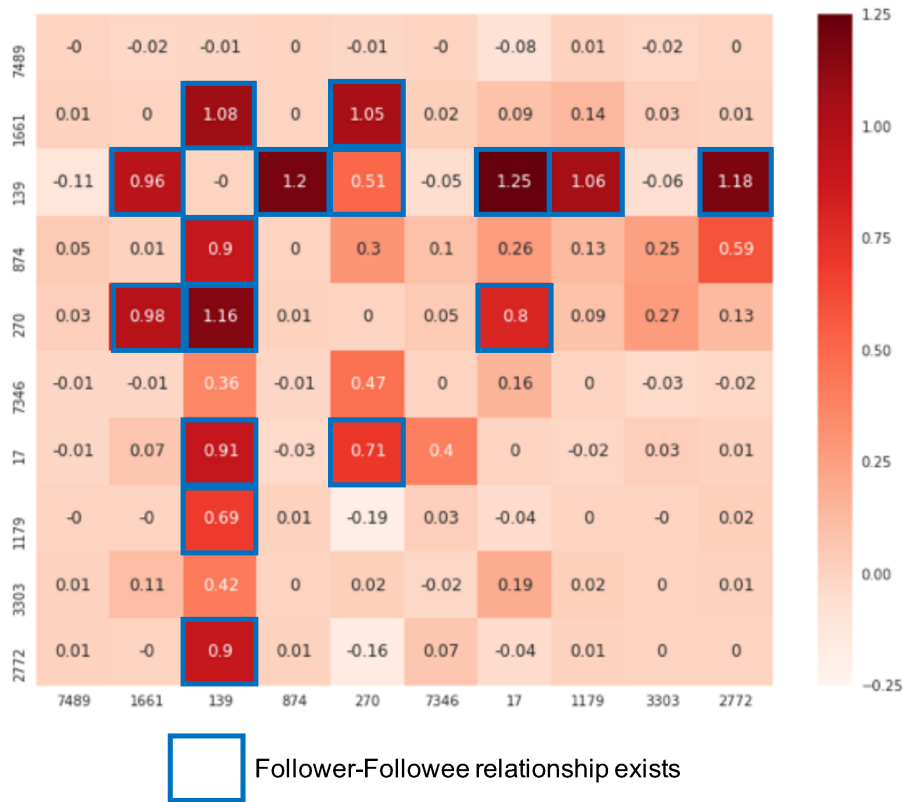


図 5.2: ビットコインの3日後のボラティリティ予測の検証実験で獲得された影響関係を表す行列 E において、サンプリングされた10アカウントの影響関係のヒートマップ。値が大きいほど影響が大きいことを示している。

ンに対して強い関心を持ち、実際にビットコインに対して投資をしていると推察される。一方、ユーザ u_{874} はツイッターで有名な暗号資産ディーラーであり、ビットコインや他の暗号資産について何度も投稿し、情報発信を行っていた。

この結果は、ソーシャルメディアのネットワークにおける直接的な関係がなくとも、お互いの興味関心と情報発信の内容が類似しているユーザは、時間を通して影響関係になる可能性を示唆している。

5.4.2 金融資産の性質に対するソーシャルメディア特徴量の予測に対する重要度の分析

実験では、暗号通貨市場を対象に検証を行ったが、ソーシャルメディアにおける影響関係が金融市場に作用するその度合いは、金融資産の性質にも依存するという仮説が成り立つ。そこで、需要と供給がより投資家の心理に左右されると推察される暗号資産だけでなく、日本の株式市場において提案手法を適用し、事業内容や市場特性ごとに最終層のパラメタ \mathbf{W}_{out} を可視化することによって、 s_t と h_t の予測に対する重要度を獲得し、各金融資産の性質に応じてソーシャルメディアにおける事象がどれだけ予測に寄与しているかを相対的に分析・考察する。

ツイッターデータは5.3節と同様のものを用いて、日本の株式市場のデータは、Yahoo Finance² から得られる2017年1月1日から2017年12月31日までの1年間のデータを利用する。価格変動を表す特徴量 m_t は、5.3節と同様に計算する。関連するソーシャルメディアへの投稿は、日本株の証券コードおよび企業名を含む投稿を対象とした。なお、式(5.6)の入力変数は、共に独立であることが証明できず、かつ標準化されていないため、パラメタの絶対値から予測に対する重要度を議論することはできないことに注意する。 \mathbf{W}_{out} におけるパラメタの初期値を全て0.5にし、日本株式の各金融資産ごとに5日後の価格変動のボラティリティ予測のための学習を行い、学習後の \mathbf{W}_{out} のパラメタを測定することによって、ソーシャルメディアにおける事象が金融市場の予測に相対的に寄与しやすい特性を分析する。

各事業・市場において、以下に示される2017年時点で上場しておりかつ時価総額上位5銘柄を選択し、分析対象とした。

- バイオ事業: サンバイオ (4592), そーせい (4565), ユーグレナ (2391), アンジェス (4563), ペプチドリーム (4587)

²<https://finance.yahoo.co.jp/>

- ソーシャルゲーム事業: ガンホー・オンライン・エンターテイメント (3765), Klab(3656), イグニス (3689), ドリコム (3793), アカツキ (3932)
- インターネット事業: ソフトバンクグループ (9984), 楽天 (4755), ヤフー (4689), サイバーエージェント (4751), ディー・エヌ・エー (2432)
- 自動車事業: トヨタ自動車 (7203), 本田自動車 (7267), 日産自動車 (7201), スズキ (7269), SUBARU(7270)
- 銀行事業: 三菱UFJファイナンシャルグループ (8306), ゆうちょ銀行 (7182), 三井住友ファイナンシャルグループ (8316), みずほファイナンシャルグループ (8411)
- 東証一部市場: トヨタ自動車 (7203), NTTドコモ (9437), 日本電信電話 (9432), ソフトバンクグループ (9984), 三菱UFJファイナンシャルグループ (8306)
- 東証二部市場: 東芝 (6502), アトム (7412), 三谷商事 (8066), 帝国ホテル (9708), 日本精機 (7287)
- マザーズ市場: サンバイオ (4592), ミクシィ(2121), ジャパンインベストメントアドバイザー (7272), ティーケーピー (3479), U U U M (3990)

上記の事業, 市場ごとに W_{out} の分析を行った結果は以下の通り.

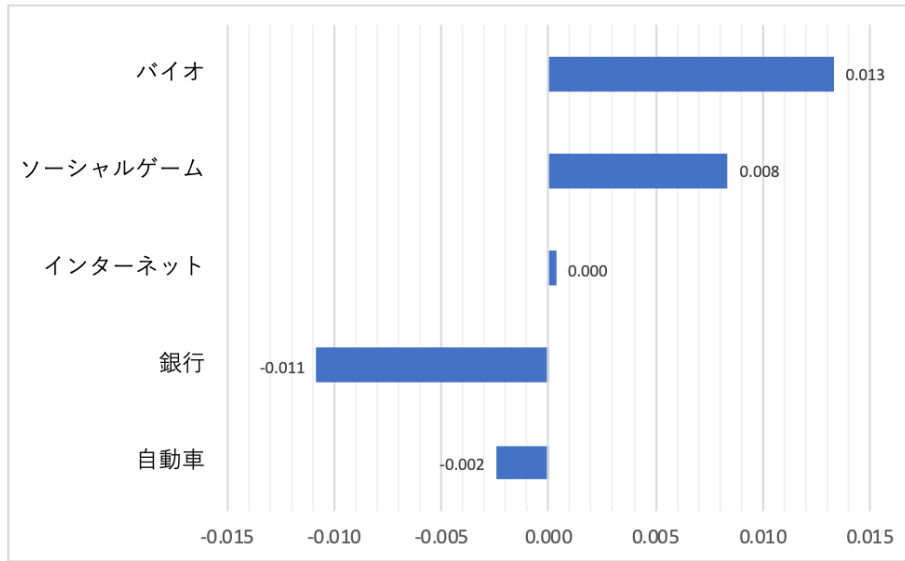


図 5.3: 事業ごとの W_{out} の分析結果. 金融市場における価格変動を表す変数 h_t への重みに対して, ソーシャルメディアにおける事象を表す変数 s_t への重みの差分を表している. 値が大きいほどソーシャルメディアが予測に寄与する比重が大きい. バイオ事業やソーシャルゲーム事業において, よりソーシャルメディアにおける事象を表す変数 s_t が予測に寄与していることがわかる.

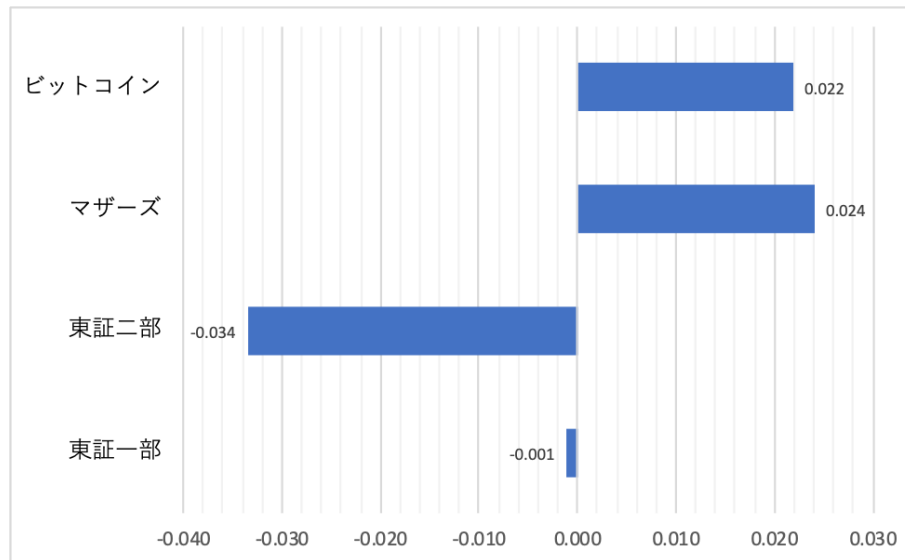


図 5.4: 市場ごとの W_{out} の分析結果. 金融市場における価格変動を表す変数 h_t への重みに対して, ソーシャルメディアにおける事象を表す変数 s_t への重みの差分を表している. ビットコイン市場やマザーズ市場において, よりソーシャルメディアにおける事象を表す変数 s_t が予測に寄与していることがわかる.

各事業・市場ごとに, 学習過程における W_{out} の推移を見てみると, 自動車関連事業および銀行関連事業はツイッターにおける事象を表す変数 s_t に対する重みよりも, 金融市場における価格変動を表す変数 h_t に対する重みが大きくなっていることがわかる. 一方で, ソーシャルゲームやバイオ関連事業は, ツイッターにおける事象を表す変数 s_t に対する重みが, 金融市場における価格変動を表す変数 h_t に対する重みよりも大きくなっていることがわかる. インターネット関連事業は, それぞれの重みはそれぞれほぼ等しかった. また, 東証1部市場の銘柄は, より金融市場における価格変動を表す変数 h_t に対する重みが大きかったのに対して, マザーズ市場やビットコインは, よりツイッターにおける事象を表す変数 s_t に対する重みが大きかった.

自動車や銀行関連銘柄, 東証一部の銘柄などの事業規模も大きい成熟産業銘柄よりも, ソーシャルゲームやバイオ関連事業, またマザーズ市場やビットコインといった, 時価総額が小さく, プレスリリースなどの1つのニュースでも価格変動

が比較的簡易に起こる金融資産は、ソーシャルメディアのユーザである個人投資家が好む金融資産である。金融資産・市場の性質によって、ソーシャルメディアから金融市場への影響の強さが異なることが示唆される。

5.5 終わりに

ソーシャルメディアは、金融市場において投資家が何を認知し、どのような感情を抱いているかを獲得することができるセンサとして機能しているだけでなく、金融市場の価格変動に働きかけるアクチュエータとしても機能している可能性がある。そこで本研究は、直接観測することができず、時間を通して相互に複雑に関連している、ソーシャルメディアにおける影響関係から金融市場への作用を、RNNを用いてモデル化する手法を提案した。

実験を通して、ソーシャルメディアにおける影響関係から金融市場への作用をモデル化する提案手法が、既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻にかつ安定して予測できることを示した。また、提案手法のモデルは、ソーシャルメディアの影響関係から金融市場への作用の分析可能性を高めた。分析の中で、ソーシャルメディアのネットワークが直接接続されていなくとも、時間を通して影響関係になる可能性があり、さらに金融資産・市場の性質によってソーシャルメディアから金融市場への影響度が異なることを示した。

なお、本研究では、ソーシャルメディアにおける影響関係から金融市場に対する作用をモデル化することで、金融資産の価格変動をより精緻にかつ安定して予測することを目的としている。ニューラルネットワークの構造や学習手法そのものは、その性能や分析可能性を考慮し、異なる手法を適用できるという点で拡張性を持ち、改良の余地がある。また、本実験のデータ期間として2017年を対象としたが、暗号資産市場は上昇局面であった。金融資産・市場の性質によってソーシャルメディアから金融市場への影響度が異なるため、様々な市場に本研究を適用す

るためには、更なる実験が必要となる可能性がある。

第6章 考察

本研究では、既存の金融理論の中でヒトの認知処理限界を理由に単純化されていた手法に対して深層学習を用いることによって、人の認知処理限界に依存しない金融理論を構築できる可能性があるという仮説を立てた。既存の金融理論に基づく金融市場の予測手法に対して、今までヒトの認知処理限界が背景にあったモデルを、深層学習を用いたアルゴリズムを組み込み拡張するための研究を行った。

金融市場の予測モデルでは、分析する対象は i) 市場, ii) 金融資産, iii) 投資家の3つに分類できる。このうち、図 1.1 で整理されるように、データ構造の画一性が低く、情報の非対称性が相対的に大きい ii) 金融資産, iii) 投資家を対象にする分析は、ヒトの認知処理限界による制約が大きいと考えた。

そこで本研究では、まず、金融資産の特徴や性質に着目して株主価値を予測する際、ヒトでは1つないしは数企業の財務モデリングを行う事が限界であった手法に対して、Deep Learning を用いて複数の企業の財務・業績情報を同時に取り扱う手法に拡張し、その有用性を検証した(研究1)。次に、ポートフォリオのリスクを予測する際、ヒトではポートフォリオを構成する全ての企業のビジネスモデルや事業内容とそのリスクの関係を知らずに限界があったのに対して、Deep Learning を用いて企業の事業内容のベクトル表現を獲得し、ビジネスモデルや事業内容を反映した共分散行列の予測手法に拡張し、その有用性を検証した(研究2)。最後に、投資家の心理や行動に着目して価格変動を予測する際、ヒトでは社会ネットワークの構造や自ら情報発信をしない投資家を分析することには限界があったのに対して、深層学習を用いてソーシャルメディア内のネットワーク構造と、直接観測することが難しい投資家の内部状態から金融市場への作用をモデル化した予測手法に拡張し、その有用性を検証した(研究3)。

本章では、これらの各研究に対する研究結果をまとめ、それぞれの研究により得られた知見について述べる。また、各研究には手法の限界や課題があるため、それらについて考察する。最後に、本研究を通して新たに生まれた「問い」に言及し、本研究をさらに発展させることで将来どのようなことが可能になるか、将来の展望を述べる。

6.1 3章から5章までのまとめ

まず、既存のテクニカル分析の機械学習手法を用いるだけでなく、市場における価値に着目したファンダメンタルズ分析を機械学習を用いて行い、アンサンブル学習を用いた複合的な株主価値推定モデルを提案した(研究1)。XBRLファイルを利用し、全ての上場企業の企業情報を取得するシステムを用いて企業情報を一元化した。推定対象を標準化された株主価値にすることで、機械学習によるファンダメンタルズ分析を可能にした。全企業の時系列のファンダメンタルズ情報を用いて教師あり機械学習を行うことで、帰納的にパターンを見つけ出し、重要な素性を抽出することで人の認知処理限界を解消した。検証実験を通じて、既存の経済学的ファンダメンタルズ分析や、過去の株価データを用いたテクニカル分析単体の手法よりも、提案手法がより正確な株主価値の推定を行えることを示した。

次に、株価情報だけを考慮するのではなく、ニューラルネットワークを用いて企業のビジネスモデルや事業内容のベクトル表現を獲得し、ポートフォリオの共分散行列を予測する手法を提案した(研究2)。企業のビジネスモデルや事業内容が記述されている投資家向け資料のデータを一元化し必要な情報を抽出したのち、ニューラルネットワークによる事業のベクトル表現を用いた共分散行列を導出することで、ヒトの認知処理限界を解消し、ポートフォリオの共分散行列を予測する際に、多種多様なビジネスモデルの全てを体系的に取り扱うことを可能にした。日本株式市場を対象に検証を行い、提案手法はより精緻に共分散行列の予測ができること、ビジネスモデルや事業内容を反映した共分散行列が、株式ポートフォ

リオの最適化に有用であることを示した。

(研究3)では、投資家心理に着目し、ヒトの認知処理能力では取り扱うことが難しかったソーシャルメディアの影響関係から金融市場への作用を、深層学習を用いてモデル化することで、価格変動をより精緻にかつ安定して予測可能な手法を提案した。金融市場の観測や予測をする際、ソーシャルメディアを実世界の射影と捉え、観測・予測に有用な情報源として用いる手法が提案されていた。(研究3)は、ソーシャルメディアが、金融市場の価格変動に働きかけるアクチュエータとしても機能しているという仮説に基づく。金融市場で観測される値を予測対象とした損失関数を持つ end-to-end の RNN モデルを学習させることで、ヒトの認知処理能力を解消し、直接観測できず、実際の金融資産の価格変動に寄与する相互作用や、時間を通して複雑に変化するユーザの認知や心理状態を抽出・分析を可能にした。金融市場の予測のタスクに対して、提案手法のモデルと、ソーシャルメディアにおける影響関係から金融市場に対する作用を反映していない複数のモデルを比較検証した。既存の価格決定要因では説明することが難しい暗号資産の価格変動をより精緻にかつ安定して予測可能であることを示した。さらに、ソーシャルメディアの影響関係から金融市場への作用の分析可能性を高めたこと、その分析の中で、ソーシャルメディア内のネットワークが直接接続されていなくとも、時間を通して影響を与える関係になることや、金融資産の性質によってソーシャルメディアから金融市場への影響の強さが異なることを示唆した。

6.2 本研究の貢献

機械学習を用いて金融市場の予測を行う研究は数多くあるが、機械学習を用いることでより精緻な予測を実現できる理由を考察し、既存の金融理論の視点から仮説を立ててアプローチする研究は相対的に少ない。2.2節でも言及した通り、過去の株価やニュース情報を用いて市場の状態に着目した予測モデルの研究が多い。深層学習は、従来の特徴量抽出器・分類器よりも優れた性能を発揮し、特に学習過程で隠れ層において入力特徴空間に依存しない特徴表現を獲得できるアルゴリ

ズムである。図 2.2 の分類の中では、時系列予測モデルの関数を、深層学習をはじめとしたより優れた機械学習のモデルに置き換える研究が多くを占めている。しかし、深層学習が優れた特徴抽出の能力を持っていたとしても、データにノイズも多く含まれ、非定常性がある金融市場においては、予測モデル全体を深層学習を用いた最先端のアルゴリズムでただ置き換えただけでは上手くいかない可能性がある。実際、機械学習を用いたモデルよりも、伝統的な統計的手法を用いたモデルの方が優れている可能性を示唆する主張も存在する [4]。

本研究の目的は、既存の金融理論に基づく金融市場の予測手法に対して、今までヒトの認知処理限界が背景にあったモデルを、深層学習を用いたアルゴリズムを組み込み拡張することであった。深層学習をヒトの認知処理能力を拡張する関数として捉え、既存の金融理論や研究を構造化し、ヒトの認知処理能力の限界がボトルネックになっている機能に対して、深層学習を適用させた。(研究1)では、金融資産の特徴や性質に着目して株主価値を予測する際、ヒトでは1つないしは数企業の財務モデリングを行う事が限界であった手法に対して、Deep Learning を用いて複数の企業の財務・業績情報を同時に取り扱う手法に拡張し、その有用性を検証した。(研究2)では、ポートフォリオのリスクを予測する際、ヒトではポートフォリオを構成する全ての企業のビジネスモデルや事業内容とそのリスクの関係を知らずに限界があったのに対して、Deep Learning を用いて企業の事業内容のベクトル表現を獲得し、ビジネスモデルや事業内容を反映した共分散行列の予測手法に拡張し、その有用性を検証した。(研究3)では、投資家の心理や行動に着目して価格変動を予測する際、ヒトでは社会ネットワークの構造や自ら情報発信をしない投資家を分析することには限界があったのに対して、深層学習を用いてソーシャルメディア内のネットワーク構造と、直接観測することが難しい投資家の内部状態から金融市場への作用をモデル化した予測手法に拡張し、その有用性を示した。

本研究の結果から、予測モデル全体を深層学習を用いた最先端のアルゴリズムでただ置き換えるだけではなく、既存の金融理論に基づく予測手法の中でヒトの

認知処理限界による課題を特定し、予測可能性の仮説に基づき深層学習を適用して既存の理論や手法を再構築することの有用性と可能性が示唆される。深層学習の研究は日進月歩で進化しているが、本研究は金融市場の予測問題においてどのように深層学習を適用させることができるのかという方向性の一つを示した点において、学術的にも実用的にも価値がある。

本研究は、既存の金融理論に基づく金融市場の予測手法に対して、今までヒトの認知処理限界が背景にあったモデルを、深層学習を用いたアルゴリズムを組み込み拡張している。そのため、既存の金融理論を用いている実現場でも、実用的に日々行われている分析やデータ処理のオペレーションを、本研究の手法を用いて拡張したり、効率化することが期待できる。例えば、(研究1)の手法を用いれば、数千とある企業の決算短信を全て目を通すことなくとも、業績の相対的な評価を瞬時に獲得したり、決算短信が発表された直後に投資判断を行うことも可能になる。(研究2)の手法を用いれば、全ての投資家向け資料を読むことなく、類似性の高い企業を発見でき、ビジネスモデルや事業内容を理解せずとも、ビジネスモデルを反映したポートフォリオを構築することが可能になると考えられる。(研究3)の手法を拡張させることで、ソーシャルメディアによって増幅する個人投資家たちの心理状態や注目銘柄を素早く抽出することが可能になると考えられる。

6.3 本研究の課題と限界

本研究には、いくつかの課題や限界が存在している。特に、提案手法の時系列普遍性と、解釈・説明性とのトレードオフの観点からそれぞれについて課題を述べる。

6.3.1 時系列普遍性の課題と限界

金融市場は非定常的であるため、金融市場の予測は難しい。例えばアノマリーを発見して予測モデルを構築し、一時的に収益を得ることはできても、安定的な収

益を得られるモデルを構築することは困難である。

(研究1)では、金融資産の特徴や性質に着目して株主価値を予測する際、ヒトでは1つないしは数企業の財務モデリングを行う事が限界であった手法に対して、Deep Learningを用いて複数の企業の財務・業績情報を同時に取り扱う手法を提案した。(研究2)では、ポートフォリオのリスクを予測する際、ヒトではポートフォリオを構成する全ての企業のビジネスモデルや事業内容とそのリスクの関係を知らることに限界があったのに対して、Deep Learningを用いて企業の事業内容のベクトル表現を獲得し、ビジネスモデルや事業内容を反映した共分散行列の予測手法を提案した。それぞれ手法を適用することの有用性は検証によって示されたが、特定の期間で学習されたモデルをどのように利用し続けていくべきかについては課題が残る。マクロ環境の変化が企業の業績に影響を与える可能性もあり、異なるマクロ環境の業績を長期的に学習したモデルよりも、特定のマクロ環境に合わせて学習期間を短くしたモデルの方が優れたパフォーマンスを示す可能性がある。また、もし上述した可能性があるのであれば、実務的には、特定の期間で学習したモデルをメタ的に評価し、複数のモデルを管理する仕組みを導入するなどといった対応が必要である。

(研究3)では、投資家の心理や行動に着目して価格変動を予測する際、深層学習を用いてソーシャルメディア内のネットワーク構造と、直接観測することが難しい投資家の内部状態から金融市場への作用をモデル化した予測手法を提案した。当研究では特定のツイッターユーザをサンプリングし予測モデルを構築したが、ユーザが他のソーシャルメディアを利用したり、または何らかの事業で対象とするソーシャルメディアの利用をやめた場合、構築したモデルの精度は劣化する可能性がある。また、新たなユーザの登録やインフルエンサーの登場などに伴い、ソーシャルメディア全体の中でサンプリングすべき対象が変化していく可能性がある。2章でも説明した通り、ユーザのサンプリングの課題を解く研究はいくつか存在しているものの、動的に変化するソーシャルメディアのネットワークの取り扱いは難しく、当研究の同一モデルを運用し続けることには限界がある。

6.3.2 解釈・説明性とのトレードオフの存在

優れた取引戦略を実現するための要素の一つとして、2.1節で示した通り、解釈・説明性の向上が挙げられる。資産運用会社や証券会社は投資家から資金を預かり資産運用を行うため、投資家に対して取引の根拠を説明する責任がある。また、上述の通り、金融市場は非定常的であるため、実用的にはモデルの改善や管理を逐次行っていく必要があるが、予測モデルの解釈性・説明性が低いと、課題の特定ができず改善サイクルを回すことが難しくなる。

しかし、深層学習は複雑な問題を複雑なまま解くことができることが長所である一方で、深層学習の関数は人が認識することのできないほどのパラメタを内部に擁しており、解釈することが難しい。本研究では、既存の金融理論に基づく金融市場の予測手法に対して、今までヒトの認知処理限界が背景にあったモデルを、深層学習を用いたアルゴリズムを組み込み拡張することを目的としていたが、それ故に、ヒトが認知処理することのできない機能を予測手法に内包することになる。予測モデルが深層学習によって高度になることと、解釈・説明性には一定のトレードオフが存在する。

6.4 新たな問いと展望

本節では、本研究を通して新たに生まれた「問い」に対して考察を行い、本研究をさらに発展させることで将来どのようなことが可能になるか、将来の展望を述べる。

6.4.1 深層学習を用いたヒトの認知処理限界の解消と金融理論の拡張可能性はどれだけあるか

深層学習を用いてヒトの認知処理限界を解消し、既存の金融理論を再構築する研究は今後も拡張可能であると考えられる。本研究の拡張可能性と方向性として、大きく以下の2つの方向性が挙げられる。

1. 本研究で取り扱った金融市場の予測モデルとは異なる切り口で既存の金融理論を切り出し、ヒトの認知処理限界による課題を深層学習によって解消する研究を広げていく方向性
2. 本研究で取り扱った金融市場の予測モデルに対して、課題の改善や適用市場の拡張を進めていく方向性

本研究では、金融市場の予測モデルを、分析する対象の i) 市場, ii) 金融資産, iii) 投資家の3つに分類した。データ構造の画一性が低く、情報の非対称性が相対的に大きい ii) 金融資産, iii) 投資家を対象にする分析がよりヒトの認知処理限界による制約が大きいと考え、(研究1), (研究2), (研究3)に取り組んだ。本研究を通して、予測モデル全体を深層学習を用いた最先端のアルゴリズムでただ置き換えるだけではなく、既存の金融理論に基づく予測手法の中でヒトの認知処理限界による課題を特定し、予測可能性の仮説に基づき深層学習を適用して既存の理論や手法を再構築することの有用性と可能性を示唆した。この示唆は、(研究1), (研究2), (研究3)で取り扱った金融市場の予測モデルや既存の金融理論にも当てはまると期待できる。本研究で取り扱った金融市場の予測モデルとは異なる切り口で既存の金融理論を切り出し、ヒトの認知処理限界による課題を深層学習によって解消する研究を広げていくことを展望できる。

本研究で取り扱った金融市場の予測モデルに対して、課題の改善や適用市場の拡張を進めていく方向性としては、6.3節で言及した時系列普遍性の課題と限界及び解釈・説明性とのトレードオフの存在を解消する追加の研究を行っていくことが望ましい。

6.4.2 深層学習によって金融市場は効率的になるのか

1.1節では、「金融市場の予測とは、投資家がもつモデルと現時点での社会的な合意との差分を認識することであり、金融市場の予測モデルとは、投資家がもつモデルと現時点での社会的な合意との認知差分を理論化したものである」と述べた。深層学習を用いてヒトの認知処理限界が解消され、既存の金融理論の再構築が進んでいくと、投資家が認知できる情報量や速度といった知識・知能差が低減し、投資家がもつモデルと現時点での社会的な合意との差分は少なくなっていくと考えられる。結果として、ヒトの認知処理限界や認知処理能力の差によって生まれていた予測可能性の余地は少なくなっていくと思料する。

その上で、市場が効率的になるためには、深層学習を用いた金融理論や予測モデルが数多くの投資家に対して利用可能になる必要がある。ヒトの認知処理限界によって予測可能性の余地があり、 α を追求することができるのであれば、自らがそのリターンを享受すればよく、基本的に他者に対して優れた金融理論や予測モデルを共有するメリットは少ない。そのため、深層学習を用いてヒトの認知処理限界が解消され、既存の金融理論の再構築が進んでいっても（または一部の投資家の中では既に進んでいたとしても）、すぐには市場は効率的になることはないと推察できる。学術的なコミュニティや、差別化に迫られる証券会社が、さらなる研究や付加価値提供として徐々に知見が共有されていく可能性が高い。

深層学習を用いた金融理論や予測モデルが数多くの投資家に対して利用可能になり、投資家が認知できる情報量や速度といった知識・知能差が低減し、投資家がもつモデルと現時点での社会的な合意との差分は少なくなっていくと仮定する。結果として、金融市場は効率的になり、 α の追求は困難になる代わりに、今まで発生していた知的労働コストや、非合理で不要な取引コストがなくなるため、市場全体が生み出す付加価値は大きくなると考えられる。

金融市場が効率的になっていき、 α の追求が難しくなると、投資家が金融市場に期待する役割が変化するのではないかと推察する。前提として、「金融市場に投資

する投資家は、最小のリスクで最大のリターンを得ることを期待する」というものがあつたが、SDGs(Sustainable Development Goalsの略、以下SDGsと呼ぶ)の考え方に代表されるように、社会の持続可能性や、既存の資本主義に捉われない新たな「意味」を市場に見出すようになるかもしれない。

第7章 結論

本論文では、既存の金融理論に基づく金融市場の予測手法に対して、今までヒトの認知処理限界が背景にあったモデルを、深層学習を用いたアルゴリズムを組み込み拡張することを目的として研究を行った。

第1章では、既存の金融理論を整理しながら金融市場の予測という問題を解釈し、ヒトの認知処理限界を理由に単純化されていた既存の金融理論の手法を拡張することができれば、より精緻な予測が可能になると仮説を立てた。深層学習は、従来の特徴量抽出器・分類器よりも優れた性能を発揮し、特に学習過程で隠れ層において入力特徴空間に依存しない特徴表現を獲得できるアルゴリズムであるが、データにノイズも多く含まれ、非定常性がある金融市場においては、予測モデル全体を深層学習でただ置き換えただけでは上手くいかない可能性について言及した。その後、金融市場の予測モデルを、分析する対象の i) 市場, ii) 金融資産, iii) 投資家の3つに分類し、データ構造の画一性が低く、情報の非対称性が相対的に大きい ii) 金融資産, iii) 投資家を対象にする分析は、ヒトの認知処理限界による制約が大きいと考えられる説明し、具体的な研究目的を設定した。

第2章では、本研究の学術的見地を明確にした。はじめに、金融市場の予測モデルの研究を、予測の目的や取り扱う情報を軸に整理し、金融市場における予測のタスクの解像度を高めた。次に、市場の状態に着目した予測モデル、金融資産の性質に着目した予測モデル、投資家に着目した予測モデルのそれぞれに対して、既存研究を整理し述べた。

第3章では、既存のテクニカル分析の機械学習手法を用いるだけでなく、市場における価値に着目したファンダメンタルズ分析を機械学習を用いて行い、アンサ

ンブル学習を用いた複合的な株主価値推定モデルを提案した。XBRL ファイルを利用し、全ての上場企業の企業情報を取得するシステムを用いて企業情報を一元化した。推定対象を標準化された株主価値にすることで、機械学習によるファンダメンタルズ分析を可能にした。全企業の時系列のファンダメンタルズ情報を用いて教師あり機械学習を行うことで、帰納的にパターンを見つけ出し、重要な素性を抽出することで人の認知処理限界を解消した。検証実験を通じて、既存の経済学的ファンダメンタルズ分析や、過去の株価データを用いたテクニカル分析単体の手法よりも、提案手法がより正確な株主価値の推定を行えることを示した。

第4章では、株価情報だけを考慮するのではなく、ニューラルネットワークを用いて企業のビジネスモデルや事業内容のベクトル表現を獲得し、ポートフォリオの共分散行列を予測する手法を提案した。企業のビジネスモデルや事業内容が記述されている投資家向け資料のデータを一元化し必要な情報を抽出したのち、ニューラルネットワークによる事業のベクトル表現を用いた共分散行列を導出することで、ヒトの認知処理限界を解消し、ポートフォリオの共分散行列を予測する際に、多種多様なビジネスモデルの全てを体系的に取り扱うことを可能にした。日本株式市場を対象に検証を行い、提案手法はより精緻に共分散行列の予測ができること、ビジネスモデルや事業内容を反映した共分散行列が、株式ポートフォリオの最適化に有用であることを示した。

第5章では、投資家心理に着目し、ヒトの認知処理能力では取り扱うことが難しかったソーシャルメディアの影響関係から金融市場への作用を、深層学習を用いてモデル化することで、価格変動をより精緻にかつ安定して予測可能な手法を提案した。金融市場で観測される値を予測対象とした損失関数を持つ end-to-end の RNN モデルを学習させることで、ヒトの認知処理能力を解消し、直接観測できず、実際の金融資産の価格変動に寄与する相互作用や、時間を通して複雑に変化するユーザの認知や心理状態を抽出・分析を可能にした。金融市場の予測のタスクに対して、提案手法のモデルと、ソーシャルメディアにおける影響関係から金融市場に対する作用を反映していない複数のモデルを比較検証した。既存の価格決定

要因では説明することが難しい暗号資産の価格変動をより精緻にかつ安定して予測可能であることを示した。さらに、ソーシャルメディアの影響関係から金融市場への作用の分析可能性を高めたこと。その分析の中で、ソーシャルメディア内のネットワークが直接接続されていなくとも、時間を通して影響を与える関係になることや、金融資産の性質によってソーシャルメディアから金融市場への影響の強さが異なることを示唆した。

第6章では、第3章から第5章までの各研究に対する研究結果をまとめ、それぞれの研究により得られた知見について述べた。各研究には手法の限界や課題があるため、それらについて考察した。本研究を通して新たに生まれた「問い」に言及し、本研究をさらに発展させることで将来どのようなことが可能になるか、将来の展望を述べた。

これらの研究を通して、予測モデル全体を深層学習を用いた最先端のアルゴリズムでただ置き換えるだけではなく、既存の金融理論に基づく予測手法の中でヒトの認知処理限界による課題を特定し、予測可能性の仮説に基づき深層学習を適用して既存の理論や手法を再構築することの有用性と可能性を示した。しかしながら、第6章で言及したように残存する課題や、新たに生まれた仮説や問いも存在している。これらの課題や問いに向き合い、さらなる研究と実践を重ねていきたい。

参考文献

- [1] Burton G Malkiel and Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, Vol. 25, No. 2, pp. 383–417, 1970.
- [2] Ashiq Ali, Lee-Seok Hwang, and Mark A Trombley. Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, Vol. 69, No. 2, pp. 355–373, 2003.
- [3] Herbert A Simon. *Administrative behavior*. Simon and Schuster, 2013.
- [4] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one*, Vol. 13, No. 3, p. e0194889, 2018.
- [5] J Bradford De Long, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. Noise trader risk in financial markets. *Journal of political Economy*, Vol. 98, No. 4, pp. 703–738, 1990.
- [6] Eugene F Fama. Random walks in stock market prices. *Financial analysts journal*, Vol. 51, No. 1, pp. 75–80, 1995.
- [7] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, Vol. 19, No. 3, pp. 425–442, 1964.
- [8] Eugene F Fama and Kenneth R French. *Common risk factors in the returns on stocks and bonds*. University of Chicago Press, 2021.

- [9] 松尾豊. ウェブからの実世界の観測と予測. 電子情報通信学会論文誌 B, Vol. 96, No. 12, pp. 1309–1315, 2013.
- [10] William Brock, Josef Lakonishok, and Blake LeBaron. Simple technical trading rules and the stochastic properties of stock returns. *The Journal of finance*, Vol. 47, No. 5, pp. 1731–1764, 1992.
- [11] Jeffrey A Frankel and Kenneth A Froot. Using survey data to test some standard propositions regarding exchange rate expectations, 1985.
- [12] Jeffrey A Frankel and Kenneth Froot. Chartists, fundamentalists, and trading in the foreign exchange market. 1990.
- [13] Rudiger Dornbusch. Expectations and exchange rate dynamics. *Journal of political Economy*, Vol. 84, No. 6, pp. 1161–1176, 1976.
- [14] Baruch Lev and S Ramu Thiagarajan. Fundamental information analysis. *Journal of Accounting research*, pp. 190–215, 1993.
- [15] John Maynard Keynes. *The general theory of employment, interest, and money*. Springer, 2018.
- [16] Robert J Shiller. Do stock prices move too much to be justified by subsequent changes in dividends?, 1980.
- [17] Rohit Choudhry and Kumkum Garg. A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, Vol. 39, No. 3, pp. 315–318, 2008.
- [18] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [19] Kai Chen, Yi Zhou, and Fangyan Dai. A lstm-based method for stock returns prediction: A case study of china stock market. In *2015 IEEE international conference on big data (big data)*, pp. 2823–2824. IEEE, 2015.

- [20] Ryo Akita, Akira Yoshihara, Takashi Matsubara, and Kuniaki Uehara. Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pp. 1–6. IEEE, 2016.
- [21] Hyun-jung Kim and Kyung-shik Shin. A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. *Applied Soft Computing*, Vol. 7, No. 2, pp. 569–576, 2007.
- [22] Lijuan Cao and Francis EH Tay. Financial forecasting using support vector machines. *Neural Computing & Applications*, Vol. 10, No. 2, pp. 184–192, 2001.
- [23] W Kreesuradej, Donald C Wunsch, and M Lane. Time delay neural network for small time series data sets. 1994.
- [24] Emad W Saad, Danil V Prokhorov, and Donald C Wunsch. Advanced neural network training methods for low false alarm stock trend prediction. In *Proceedings of International Conference on Neural Networks (ICNN'96)*, Vol. 4, pp. 2021–2026. IEEE, 1996.
- [25] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 95–104, 2018.
- [26] Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, Vol. 90, p. 106181, 2020.
- [27] Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 2141–2149, 2017.

- [28] George EP Box and David A Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, Vol. 65, No. 332, pp. 1509–1526, 1970.
- [29] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [30] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.
- [31] Kei Nakagawa, Takumi Uchida, and Tomohisa Aoshima. Deep factor model. In *ECML PKDD 2018 Workshops*, pp. 37–50. Springer, 2018.
- [32] David Blitz, Joop Huij, and Martin Martens. Residual momentum. *Journal of Empirical Finance*, Vol. 18, No. 3, pp. 506–521, 2011.
- [33] Kentaro Imajo, Kentaro Minami, Katsuya Ito, Kei Nakagawa, et al. Deep portfolio optimization via distributional prediction of residual factors. *arXiv preprint arXiv:2012.07245*, 2020.
- [34] Fama Eugene and Kenneth French. The cross-section of expected stock returns. *Journal of Finance*, Vol. 47, No. 2, pp. 427–465, 1992.
- [35] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.
- [36] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pp. 261–269, 2018.
- [37] Manuel R Vargas, Beatriz SLP De Lima, and Alexandre G Evsukoff. Deep learning for stock market prediction from financial news articles. In *2017*

- IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*, pp. 60–65. IEEE, 2017.
- [38] Guosheng Hu, Yuxin Hu, Kai Yang, Zehao Yu, Flood Sung, Zhihong Zhang, Fei Xie, Jianguo Liu, Neil Robertson, Timpathy Hospedales, et al. Deep stock representation learning: From candlestick charts to investment decisions. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 2706–2710. IEEE, 2018.
- [39] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, Vol. 10, No. 5, pp. 603–621, 2003.
- [40] Jianqing Fan, Yingying Fan, and Jinchi Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, Vol. 147, No. 1, pp. 186–197, 2008.
- [41] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pp. 593–607. Springer, 2018.
- [42] Yingmei Chen, Zhongyu Wei, and Xuanjing Huang. Incorporating corporation relationship via graph convolutional neural networks for stock price prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 1655–1658, 2018.
- [43] Harry M Markowitz. *Portfolio selection*. Yale university press, 1968.
- [44] Yves Choueifaty and Yves Coignard. Toward maximum diversification. *The Journal of Portfolio Management*, Vol. 35, No. 1, pp. 40–51, 2008.

- [45] Roger Clarke, Harindra De Silva, and Steven Thorley. Minimum-variance portfolio composition. *The Journal of Portfolio Management*, Vol. 37, No. 2, pp. 31–45, 2011.
- [46] Kei Nakagawa, Mitsuyoshi Imamura, and Kenichi Yoshida. Risk-based portfolios with large dynamic covariance matrices. *International Journal of Financial Studies*, Vol. 6, No. 2, p. 52, 2018.
- [47] Sébastien Maillard, Thierry Roncalli, and Jérôme Teiletche. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, Vol. 36, No. 4, pp. 60–70, 2010.
- [48] Kei Nakagawa and Katsuya Ito. Taming tail risk: Regularized multiple β worst-case cvar portfolio. *Symmetry*, Vol. 13, No. 6, p. 922, 2021.
- [49] Raymond J Dolan. Emotion, cognition, and behavior. *science*, Vol. 298, No. 5596, pp. 1191–1194, 2002.
- [50] Sean Spence. Descartes’ error: Emotion, reason and the human brain. *BMJ*, Vol. 310, No. 6988, p. 1213, 1995.
- [51] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pp. 99–127. World Scientific, 2013.
- [52] John R Nofsinger. Social mood and financial economics. *The Journal of Behavioral Finance*, Vol. 6, No. 3, pp. 144–160, 2005.
- [53] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, Vol. 2, No. 1, pp. 1–8, 2011.
- [54] Xue Zhang, Hauke Fuehres, and Peter A Gloor. Predicting stock market indicators through twitter “ i hope it is not as bad as i fear ” . *Procedia-Social and Behavioral Sciences*, Vol. 26, pp. 55–62, 2011.

- [55] Michael Nofer and Oliver Hinz. Using twitter to predict the stock market. *Business & Information Systems Engineering*, Vol. 57, No. 4, pp. 229–242, 2015.
- [56] Ladislav Kristoufek. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, Vol. 3, p. 3415, 2013.
- [57] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [58] Myron J Gordon. *The Investment, Financing, and Valuation of the Corporation*. RD Irwin, 1962.
- [59] Paul R Krugman. *International Economics: Theory and Policy, 8/E*. Pearson Education India, 2008.
- [60] Kenneth A Reinert, Ramkishen S Rajan, Amy Joycelyn Glass, and Lewis S Davis. *The Princeton Encyclopedia of the World Economy(Two volume set)*, Vol. 1. Princeton University Press, 2009.
- [61] Maurice D Levi. *International Finance: Contemporary Issues*. Routledge, 2007.
- [62] R. C. Feenstra and A. M. Taylor. *International Macroeconomics*. Worth Publishers, 2008.
- [63] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860, . ACM, 2010.
- [64] Jure Leskovec and Eric Horvitz. Planetary-scale views on a large instant-messaging network. In *Proceedings of the 17th international conference on World Wide Web*, pp. 915–924. ACM, 2008.

- [65] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 33–42. ACM, 2012.
- [66] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pp. 591–600. ACM, 2010.
- [67] Emilio Ferrara, Roberto Interdonato, and Andrea Tagarelli. Online popularity and topical interests through the lens of instagram. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, pp. 24–34, . ACM, 2014.
- [68] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welp. Predicting elections with twitter: What 140 characters reveal about political sentiment. *The International Conference on Weblogs and Social Media*, Vol. 10, No. 1, pp. 178–185, 2010.
- [69] Sitaram Asur and Bernardo A Huberman. Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Volume 01, pp. 492–499, . IEEE Computer Society, 2010.
- [70] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, pp. 115–120, . Association for Computational Linguistics, 2012.
- [71] Marco Brambilla, Stefano Ceri, Emanuele Della Valle, Riccardo Volonterio, and Felix Xavier Acero Salazar. Extracting emerging knowledge from social media. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 795–804, . International World Wide Web Conferences Steering Committee, 2017.

- [72] Minglai Shao, Jianxin Li, Feng Chen, Hongyi Huang, Shuai Zhang, and Xunxun Chen. An efficient approach to event detection and forecasting in dynamic multivariate social media networks. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 1631–1639, . International World Wide Web Conferences Steering Committee, 2017.
- [73] Yupeng Gu, Yizhou Sun, Yanen Li, and Yang Yang. Rare: Social rank regulated large-scale network embedding. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pp. 359–368, . International World Wide Web Conferences Steering Committee, 2018.
- [74] Shoko Oda. The mysterious twitter user drawing a swarm of japan traders. <https://www.bloomberg.com/news/articles/2018-01-14/the-mysterious-twitter-user-attracting-a-swarm-of-japan-traders>, 2018.
- [75] Daniel M Romero, Brian Uzzi, and Jon Kleinberg. Social networks under stress. In *Proceedings of the 25th International Conference on World Wide Web*, pp. 9–20, . International World Wide Web Conferences Steering Committee, 2016.
- [76] Munmun De Choudhury. Modeling and predicting group activity over time in online social media. In *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia*, pp. 349–350, . ACM, 2009.
- [77] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, . ACM, 2003.
- [78] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. *Social Media Mining: An Introduction*. Cambridge University Press, 2014.
- [79] Lisa R Anderson and Charles A Holt. Information cascades in the laboratory. *The American economic review*, pp. 847–862, 1997.

- [80] David S Scharfstein and Jeremy C Stein. Herd behavior and investment. *The American Economic Review*, pp. 465–479, 1990.
- [81] Sushil Bikhchandani and Sunil Sharma. Herd behavior in financial markets. *IMF Staff papers*, Vol. 47, No. 3, pp. 279–310, 2000.
- [82] Eric C Chang, Joseph W Cheng, and Ajay Khorana. An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, Vol. 24, No. 10, pp. 1651–1679, 2000.
- [83] Rama Cont and Jean-Philippe Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic dynamics*, Vol. 4, No. 2, pp. 170–196, 2000.
- [84] Simon Bourigault, Sylvain Lamprier, and Patrick Gallinari. Representation learning for information diffusion through social networks: An embedded cascade model. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*, pp. 573–582, . ACM, 2016.
- [85] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1067–1077. International World Wide Web Conferences Steering Committee, 2015.
- [86] Jian Tang, Jingzhou Liu, Ming Zhang, and Qiaozhu Mei. Visualizing large-scale and high-dimensional data. In *Proceedings of the 25th International Conference on World Wide Web*, pp. 287–297. International World Wide Web Conferences Steering Committee, 2016.
- [87] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web*, pp. 925–936. ACM, 2014.
- [88] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. Deepcas: An end-to-end predictor of information cascades. In *Proceedings of the 26th Interna-*

- tional Conference on World Wide Web*, pp. 577–586, . International World Wide Web Conferences Steering Committee, 2017.
- [89] Karthik Subbian, B Aditya Prakash, and Lada Adamic. Detecting large reshare cascades in social networks. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 597–605. International World Wide Web Conferences Steering Committee, 2017.
- [90] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Structural trend analysis for online social networks. *Proceedings of the VLDB Endowment*, Vol. 4, No. 10, pp. 646–656, 2011.
- [91] Qi Cao, Huawei Shen, Keting Cen, Wentao Ouyang, and Xueqi Cheng. Deep-hawkes: Bridging the gap between prediction and understanding of information cascades. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1149–1158, . ACM, 2017.
- [92] Jia Wang, Vincent W Zheng, Zemin Liu, and Kevin Chen-Chuan Chang. Topological recurrent neural network for diffusion prediction. In *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 475–484, . IEEE, 2017.
- [93] Everett M Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.
- [94] Vijay Mahajan and Robert A Peterson. *Models for innovation diffusion*, Vol. 48. Sage, 1985.
- [95] William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proc. R. Soc. Lond. A*, Vol. 138, No. 834, pp. 55–83, 1932.
- [96] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, Vol. 58, No. 1, pp. 83–90, 1971.

- [97] Marian-Andrei RizoIU, Swapnil Mishra, Quyu Kong, Mark Carman, and Lexing Xie. Sir-hawkes: Linking epidemic models and hawkes processes to model diffusions in finite populations. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pp. 419–428. International World Wide Web Conferences Steering Committee, 2018.
- [98] 大原剛三, 齊藤和巳, 木村昌弘, 元田浩. 情報拡散モデルに基づく社会ネットワーク上の影響度分析. 2015.
- [99] ジョン J. マーフィー. 先物市場のテクニカル分析, p. 5. 一般財団法人 金融財政事情研究会, 1990.
- [100] 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. <http://www.nikkei.com/article/DGXLZ010041570Y6A121C1EN1000/>.
- [101] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *arXiv preprint arXiv:1603.02754*, 2016.
- [102] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, Vol. 13, No. 2, 2012.
- [103] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, Vol. 4, No. 2, 2012.
- [104] Wei Pan, Jide Li, and Xiaoqiang Li. Portfolio learning based on deep learning. *Future Internet*, Vol. 12, No. 11, p. 202, 2020.
- [105] Thirumalaisamy P Velavan and Christian G Meyer. The covid-19 epidemic. *Tropical medicine & international health*, Vol. 25, No. 3, p. 278, 2020.
- [106] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pp. 1188–1196. PMLR, 2014.

- [107] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.
- [108] Sanjeev Arora, Yingyu Liang, and Tengyu Ma. A simple but tough-to-beat baseline for sentence embeddings. 2016.
- [109] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [110] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135–146, 2017.
- [111] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.
- [112] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [113] Raymond Kan and Guofu Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, Vol. 42, No. 3, pp. 621–656, 2007.
- [114] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, Vol. 41, No. 6, pp. 391–407, 1990.
- [115] Dawen Liang, Laurent Charlin, James McInerney, and David M Blei. Modeling user exposure in recommendation. In *Proceedings of the 25th International Conference on World Wide Web*, pp. 951–961, . International World Wide Web Conferences Steering Committee, 2016.

-
- [116] Andriy Mnih and Ruslan R Salakhutdinov. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, pp. 1257–1264, 2008.
- [117] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958, 2014.
- [118] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

謝辞

本研究の遂行や本論文の執筆にあたり，非常に多くの方からご指導，ご支援をいただきました。心より御礼申し上げます。

指導教官である東京大学大学院工学系研究科 技術経営戦略学専攻 松尾 豊 教授には，研究構想や本論文の論理構成について，貴重なご指導をいただきました。ここに，深く謝意を表します。また，研究内容に関してご指導いただいた内容だけでなく，アカデミアの価値を社会に繋げる理念を掲げ，研究者として世の中に発信していく姿勢，教育者として若い人材を育てていく姿勢，投資家として起業家を支援する姿勢から多くのことを感じ，学びました。私は，修士課程の際に，株式会社 ACES を創業し今に至りますが，普通であれば博士課程を過ごしながら起業をすることなど叶わなかったと考えております。松尾 豊 教授の多大なるサポートと，私がこれほどまでに自由に活動することを認めてくださるその寛容さによって，現在の私があると考えております。心より感謝いたします。

本博士論文の副査をお引き受けいただきました東京大学大学院工学系研究科 技術経営戦略学専攻 吉田 好邦 教授，東京大学大学院工学系研究科 システム創成学専攻 和泉 潔 教授，東京大学大学院 経済学研究科 柳川 範之 教授，東京大学大学院工学系研究科 技術経営戦略学専攻 武田 史子 准教授には，貴重なお時間を割いて本論文の審査を行なっていただき，誠にありがとうございました。この場を借りて御礼申し上げます。

本論文のそれぞれの研究について、上野山 勝也 元助教授，大澤 昇平 元助教授，岩澤 有祐 講師には，研究の方向性や論文の構成，さらには適語表現など細部に渡り，多大なご指導をいただきました。岩澤 有祐 講師，鈴木 雅大 特任教授，落合 桂一 特任教授には，休日にも関わらず，社会人博士課程向けの研究会を実施していただき，誠にありがとうございます。絶対的な時間投下量を確保できない中，皆様の適切かつ本質的な問いや議論，ご指導をいただいたことで，研究を前に進めることができました。深く感謝致します。

東京大学大学院工学系研究科 技術経営戦略学専攻 丸岡 賢人氏は，本論文の一部研究について，データベースの整備やライブラリの実装を手伝っていただきました。実験の速度が向上し，素早い仮説検証が実現可能になったことで，本研究を遂行することができました。丸岡氏がいなければ，半年から1年ほどさらに時間がかかっていたかもしれません。誠にありがとうございます。

松尾研究室の皆様には，多大なご協力，ご支援いただきました。学術支援職員である皆様には，分析サーバやGPU解析環境の構築，秘書業務など，日頃から研究室の環境を整えて下さり，研究生活を支えてくださいました。深く感謝致します。研究室同期である阿久澤 圭氏，小林 由弥氏，中川 大海氏は，互いに切磋琢磨し合いながら研究を進めさせていただきました。皆が論文を執筆し発表する姿を見て，自分も頑張らねばならないという気持ちにさせてくれました。ありがとうございます。ここに，松尾研究室の皆様へ謝意を表します。

株式会社 ACES の皆様方には，創業当初から多大なるサポートをいただきました。特に創業メンバーである與島 仙太郎氏，中川 大海氏，三田村 健氏，西條 真史氏，久保 静真氏は，苦楽を共に過ごしましたし，私が事業に取り組みながらも研究することをいつでも支えてくれました。本当にありがとうございます。これからは，事業を通じて，アカデミアの価値を社会に繋げることが私の社会における役割の一つであると考えています。引き続きよろしく申し上げます。

最後に、これまで支えてくれた家族、友人、パートナーの皆様へ感謝します。皆様のおかげで、心身ともに健康で研究活動に取り組み、3年間を楽しく過ごすことができました。

東京大学工学系研究科 技術経営戦略学専攻

松尾研究室 博士3年

田村浩一郎

令和4年3月

発表文献

学会誌論文

- 田村浩一郎, 上野山勝也, 飯塚修平, 松尾豊: 深層学習を用いたアンサンブルモデルによる株主価値推定モデルの提案, 人工知能学会論文誌, Vol. 33, No. 1 (2017)
- 田村 浩一郎, 松尾 豊: ソーシャルメディアにおける影響関係から金融市場に対する作用のモデル化と分析, 人工知能学会論文誌, Vol.35, No.6 (2020)

学会発表

- 田村 浩一郎, 松尾 豊, 上野山 勝也: 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, 人工知能学会全国大会 2017, 愛知県名古屋市 (ウィンクあいち)
- 田村 浩一郎, 大澤 昇平, 松尾 豊: Twitter から仮想通貨市場に対する作用の分析, 人工知能学会全国大会 2018, 鹿児島県鹿児島市

講演

- 田村 浩一郎: セミナー (研究I部): 『AI (人工知能) ×ファンダメンタルズ分析×テクニカル分析×クオンツ分析×実際のマーケット』, 特定非営利活動法人日本テクニカルアナリスト協会, 2017
- 田村 浩一郎: AI・データサイエンス概論「第3章: 画像・映像・3次元データを扱う AI の導入」, 日本オープンオンライン教育推進協議会, 2020
- 田村 浩一郎: AI・データサイエンス専門I「第4章: 画像理解と深層学習」, 日本オープンオンライン教育推進協議会, 2020