

可読性の高い解釈可能なモデルによるAIトレーダーの説明

株式会社かんぽ生命保険
運用企画部 運用計画担当
石原 龍太*

1 はじめに

深層学習モデルをはじめとするAI技術を用いたモデルには“内部が複雑なブラックボックスであるがゆえ、モデルが最適とした判断・行動の理由を人間が容易に理解できない”という解釈性¹の問題が指摘されている。例えば「資産運用にAI技術を用いた取引戦略モデルを利用する場合において、モデルの下した投資判断の妥当性を人間が容易に検証できない」といったものである。

この問題はAI技術を用いたモデルを利用する上で避けられないものであるが、モデル内部の処理プロセスの難解さはモデルの頑健性や経済的合理性に疑問をもたらすことになる。特に、金融市場のような観測データの構造が短期間で変化する可能性のある事象を対象とするモデルにおいては、精度と解釈性のバランスを追求することが実務上の有益さに資するものと考えられる。

AI技術を用いたモデル内部の処理プロセスを説明する代表的な方法として、原(2018)は以下の4つを紹介している。

(1) 大域的な説明

複雑なブラックボックスモデルを、可読性の高い解釈可能なモデルで表現することで説明する方法。

(2) 局所的な説明

特定の入力に対するブラックボックスモデルの予測の根拠を提示することで説明する方法。

(3) 説明可能なモデルの設計

最初から可読性の高い解釈可能なモデル（構造が複雑でないモデル）をつくってしまう方法。

(4) 深層学習モデルの説明

深層学習モデル、特に画像認識モデルの説明法（モデルが画像内のどの部分を認識しているかを特定しハイライトする方法）。アプローチとしては(2)の局所的な説明に該当。

筆者は昨年、感度分析によりモデルの入力と出力の関係性を表現すること（「(2)局所的な説明」の方法）で、AI技術を用いた取引戦略モデル（以下、AIトレーダー）の投資判断プロセスの説明を試み、その成果を発表した（石原(2020)）。しかし、この方法ではモデルの入力と出力の関係性を提示することはできるものの、その間に存在する因果関係を理論立てて説明することは難しい。

* 連絡先：ryouta.ishihara.hw@jp-life.jp 本研究の内容は、筆者個人に属するものであり、筆者の所属組織の公式見解を示すものではありません。また、本研究に含まれる誤りは全て筆者の責に帰するものです。

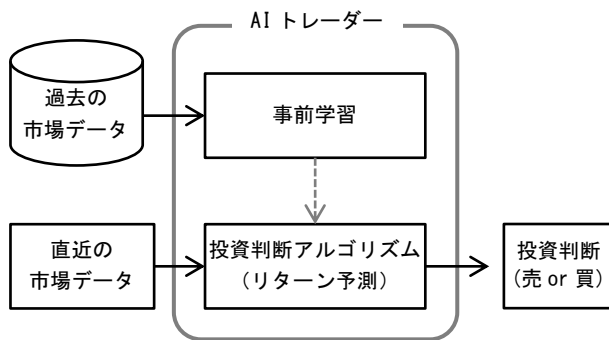
¹ 解釈性とは「モデルによる推論結果の因由を、人間が理解できる度合い」を意味する。

そこで本研究では、AIトレーダーを可読性の高い解釈可能なモデルで表現すること（「(1)大域的な説明」の方法）で、その投資判断プロセスの説明を試みる。具体的には、AIトレーダーの投資判断プロセスを近似的に表現する“可読性の高いモデル”を遺伝的アルゴリズムの手法を用いて作成し、作成したモデルの適合率と再現率の値から、その説明能力を定量的に評価する。

2 解釈対象とするAIトレーダー

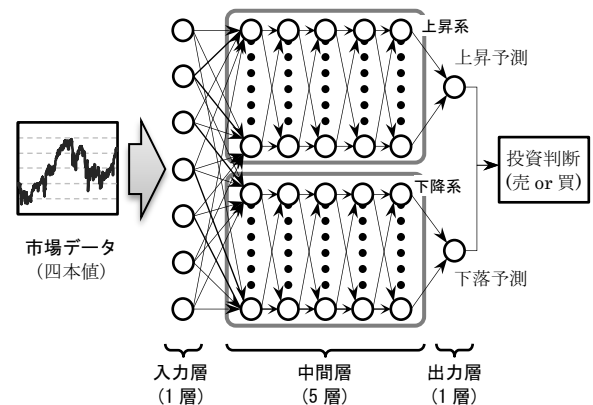
石原(2020)の研究で構築したAIトレーダー²を本研究の解釈対象とする。図1にAIトレーダーの概念図を示す。このAIトレーダーは、7層（入力層が1層、中間層が5層、出力層が1層）からなる順伝播型のニューラルネットワークを用いて構築された所謂ブラックボックスモデルであり、直近の市場データから翌日の日経225先物価格の変動（上昇または下落）を予測し、投資判断を出力する（図2）。また、過去の市場データを事前学習することで、予測精度を有するニューラルネットワークを（人手によるチューニングを必要とせず）経験的に獲得することができる。なお、AIトレーダーの詳細についての説明は、石原(2020)を参照されたい。

図1 AIトレーダーの概念図



(出所)石原 (2020)

図2 AIトレーダーの投資判断アルゴリズム



(注) 丸はノード、矢印はノード間の結合と処理の流れを示す。

(出所)石原(2020)

² 具体的には、石原(2020)において、2007年10月～2016年9月の市場データ（以下、訓練データ）を事前学習した5個体のうちの1つ（AIトレーダー1）を、本研究の解釈対象とする。

3 近似モデルの作成

3.1 近似モデルの概要

前章のAIトレーダーの投資判断プロセスを、可読性の高いモデル（以下、近似モデル）で近似的に表現し、その説明を試みる。具体的には、AIトレーダーの投資判断プロセスを日経225先物の日次収益率とボラティリティを用いて以下のように定式化する。

$$Z'_t = f(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$x = \omega_1 \text{NKf}r_t + \omega_2 \log(\sigma_t^{\text{HL}} / \overline{\sigma^{\text{HL}}}) + \lambda$$

ここで、 Z'_t は近似モデルの投資判断（買い：1，売り：0）、 $\text{NKf}r_t$ は時点 t における日経225先物の日次収益率、 σ_t^{HL} はHL法による時点 t における日経225先物の2期間のボラティリティ³、 $\overline{\sigma^{\text{HL}}}$ はボラティリティ σ_t^{HL} の期待値を表すものとする。

3.2 モデルパラメータの推計

(1) 推計方法

遺伝的アルゴリズムの手法を用いて近似モデルのパラメータ（ ω_1 , ω_2 , λ ）を推計する。本研究における遺伝的アルゴリズムの手順は次のとおりとする。①まず初期個体として、異なる遺伝子型（各個体の遺伝子は近似モデルのパラメータを意味する）を持つ個体を500個体生成させる。なお、本研究の遺伝子のとりうる値の範囲⁴は、 ω_1 が $[-1, +1]$ 、 ω_2 が $[-0.0368, +0.0368]$ 、 λ が $[-0.0177, +0.0177]$ に設定する。②次に、AIトレーダーの訓練データ（2007年10月～2016年9月の市場データ）から近似モデルの説明変数（説明変数の基本統計量は表1のとおり）を生成し、各個体の適合度を計算する。③次に、適合度の高い個体が生き残るよう遺伝的操作⁵を行う。④こうして作られた次世代についても同じ処理（適合度の計算と遺伝的操作）を10,000回繰り返す。⑤次に、遺伝的操作回数が10, 20, …, 10000回の各時点における最良個体（適合度が最も高かった個体）の中から、適合度が最も高い個体の遺伝子型を得る。⑥①～⑤の処理を50回繰り返して得られた、50通りの遺伝子型の平均を近似モデルのパラメータとする。

³ σ_t^{HL} は、以下の式により計算する。なお、 H_t は時点 t の日経225先物の高値、 L_t は時点 t の日経225先物の安値を意味する。なお、HL法についての詳細は中窪(2002)を参照されたい。

$$\sigma_t^{\text{HL}} = \sqrt{(S_t^2 + S_{t-1}^2)/2} \quad S_t^2 = 0.361 \times \{\log(H_t/L_t)\}^2$$

⁴ ω_2 の0.0368は訓練データにおける「 $\text{NKf}r_t$ の標準偏差 / $\log(\sigma_t^{\text{HL}} / \overline{\sigma^{\text{HL}}})$ の標準偏差」の値、 λ の0.0177は訓練データにおける「 $\text{NKf}r_t$ の標準偏差」の値。近似モデルの各項の値に極端な偏りが生じないように考慮して範囲を設定した。

⁵ 遺伝的操作におけるハイパーパラメータは、選択の方法をエリート選択、交叉の方法を一様交叉、突然変異の方法を置換、突然変異率を3%とする。なお、初期個体の遺伝子及び突然変異で置換する遺伝子の値は一様乱数で決定する。また、②及び④の適合度計算においては、確率的に選別した時点 t （時点 t の選別は3%の確率で行う）におけるモデルの出力値を反転（0→1, 1→0）させる（ドロップアウト）。

表1 説明変数の基本統計量（訓練データ期間）

	$NKfr_t$	σ_t^{HL}	$\log(\sigma_t^{HL}/\overline{\sigma^{HL}})$
平均	0.00%	0.93%	-13.10%
標準偏差	1.77%	0.60%	48.17%
中央値	0.09%	0.79%	-16.92%
最大値	18.81%	7.36%	206.35%
最小値	-14.00%	0.24%	-135.59%

(2) 適合度の計算

遺伝的アルゴリズムの適合度は、以下の式により計算する。

$$\text{適合度} = \frac{\sqrt{\sum_{t=1}^N Z_t \times Z'_t}}{\sqrt[4]{N^{up} \times N'^{up}}} \times \frac{\sqrt{\sum_{t=1}^N (1 - Z_t) \times (1 - Z'_t)}}{\sqrt[4]{N^{down} \times N'^{down}}}$$

このとき、 Z_t はAIトレーダーの投資判断（買い：1，売り：0）， Z'_t は近似モデルの投資判断， N^{up} はAIトレーダーの投資判断が“買い”である回数， N^{down} はAIトレーダーの投資判断が“売り”である回数， N'^{up} は近似モデルの投資判断が“買い”である回数， N'^{down} は近似モデルの投資判断が“売り”である回数を表す。なお， $\sum_{t=1}^N Z_t \times Z'_t$ はAIトレーダーと近似モデルの投資判断が両方とも“買い”である回数，また， $\sum_{t=1}^N (1 - Z_t) \times (1 - Z'_t)$ はAIトレーダーと近似モデルの投資判断が両方とも“売り”である回数を意味する。

(3) 推計結果及びその解釈

近似モデルのパラメータの推計結果を表2に示す。日次収益率の係数である ω_1 の値とボラティリティの係数である ω_2 の値がともにマイナスであったことから、AIトレーダーの投資判断には以下のような特徴があると読み解くことができる。

特徴1 直近の日次収益率がプラスである場合には“売り”，マイナスである場合には“買い”を行う（リターン・リバーサル戦略）。

特徴2 直近のボラティリティが期待値より高い場合には“売り”，低い場合には“買い”を行う（高ボラティリティ局面でリスクを回避）。

表2 モデルパラメータの推計結果

	ω_1	ω_2	λ
近似モデル	-0.7552 (0.1839)	-0.0038 (0.0019)	0.0016 (0.0011)

(注) 括弧内の数値は、モデルパラメータ推計の際に生成した遺伝子型(50通り)の標準偏差を表す。

4 近似モデルの評価

4.1 近似モデルの説明能力

前章で作成した近似モデルの説明能力を、適合率 (Precision) と再現率 (Recall) の値を用いて評価する。表3に訓練データ期間における近似モデルの適合率と再現率を示す。表3から、再現率は“売り”、“買い”ともに60%を超える水準にあることがわかる。再現率は、近似モデルがAIトレーダーの投資判断をどれだけカバーできているか、という割合 (カバー率) を表しているため、近似モデルは「AIトレーダーの投資判断をある程度表現できている」といえる。一方、適合率は“売り”が総じて低い水準にあることがわかる。売りの適合率は、近似モデルが“売り”と判断した局面のうち、AIトレーダーも“売り”と判断していた割合 (正解率) を表しているため、近似モデルには「AIトレーダーが“買い”と判断した局面を“売り”として誤検出してしまう傾向がある」といえる。これらのことから、近似モデルはAIトレーダーの投資判断をある程度表現できているものの、その説明能力は高いとは言い難い。

表3 近似モデルの適合率と再現率 (訓練データ期間)

	適合率 (Precision)			再現率 (Recall)		
	買い	売り	全体	買い	売り	全体
訓練データ期間	86.31%	34.04%	54.21%	65.91%	62.73%	64.30%
2007年10月～2010年9月	85.13%	31.55%	51.82%	62.06%	61.73%	61.90%
2010年10月～2013年9月	87.14%	36.36%	56.29%	68.35%	64.20%	66.24%
2013年10月～2016年9月	86.58%	34.49%	54.65%	67.30%	62.26%	64.74%

(注) 全体は“買い”と“売り”の幾何平均を表す。

4.2 投資パフォーマンス

次に、訓練データ期間における近似モデルの投資パフォーマンスを計測し、AIトレーダーと比較する。収益率の推移を図3、投資パフォーマンスを表4に示す。図3及び表4から、近似モデルの収益率は概ねAIトレーダーと近い動きをしているものの、総じてAIトレーダーに劣後していることがわかる。また、

表4 訓練データ期間における投資パフォーマンス

	収益率 (年率)			α (年率)		β	
	AIトレーダー	近似モデル	日経225先物	AIトレーダー	近似モデル	AIトレーダー	近似モデル
訓練データ期間	19.04%	6.26%	1.11%	18.16%	5.68%	0.7928	0.5212
2007年10月～2010年9月	13.57%	-0.09%	-18.47%	28.48%	10.43%	0.8067	0.5693
2010年10月～2013年9月	32.84%	17.08%	16.15%	19.91%	8.00%	0.8011	0.5624
2013年10月～2016年9月	10.49%	1.56%	5.66%	5.92%	-0.59%	0.8087	0.3797

図3 収益率の推移（訓練データ期間）

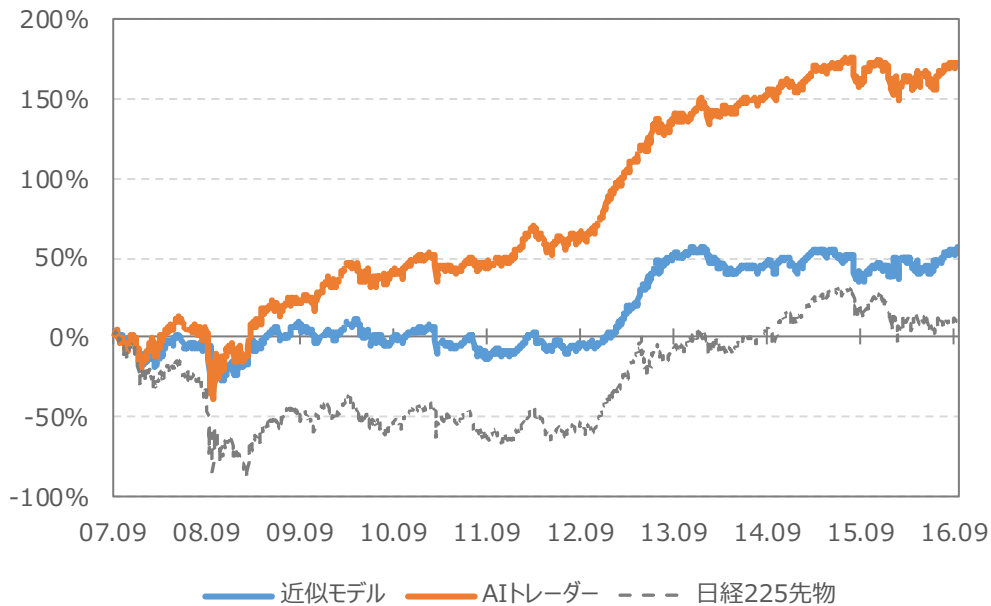


表4から近似モデルの α も総じてAIトレーダーに劣後しており、 β はAIトレーダーに比べ低い水準であることがわかる⁶。これらの結果は「近似モデルはAIトレーダーの投資判断をある程度表現できているものの、その説明能力は高いとは言い難い」という前項の考察を裏付けるものといえる。

4.3 未知データに対する投資シミュレーション

次に、2016年10月～2020年12月の市場データ（以下、外挿データ）を用いて、未知データに対する投資シミュレーションを行う。外挿データ期間における近似モデルの適合率と再現率を表5、収益率の推移を図4、投資パフォーマンスを表6に示す。外挿データ期間における近似モデルの説明能力や投資パフォーマンスは、訓練データ期間と同様の傾向を示したことから、近似モデルは未知データに対しても有効に機能しているといえる。

表5 近似モデルの適合率と再現率（外挿データ期間）

	適合率(Precision)			再現率(Recall)		
	買い	売り	全体	買い	売り	全体
外挿データ期間	93.18%	31.66%	54.31%	75.42%	67.33%	71.26%
2016年10月～2019年9月	93.10%	33.81%	56.11%	77.26%	66.36%	71.83%
2019年10月～2020年12月	93.37%	27.52%	50.69%	69.85%	69.77%	69.81%

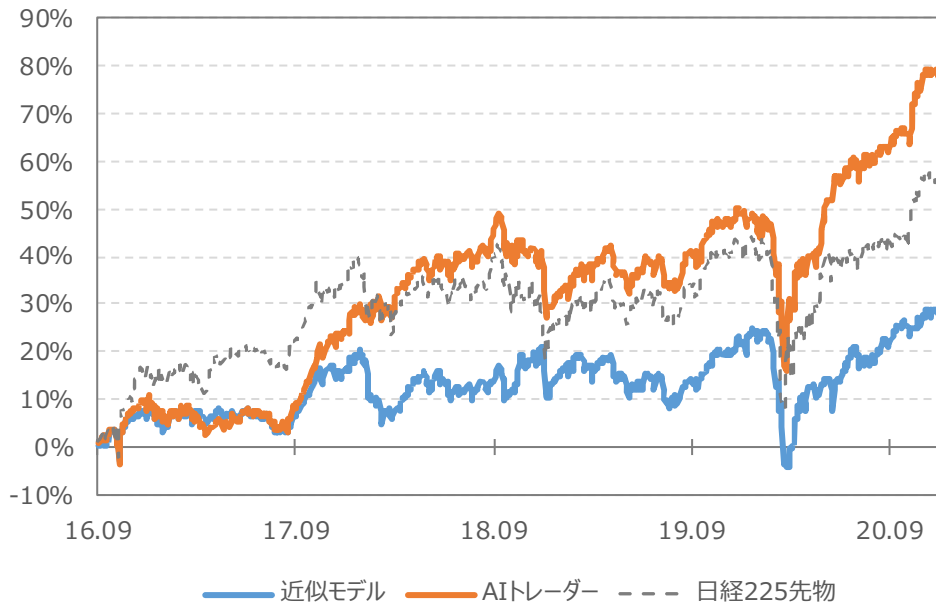
(注) 全体は“買い”と“売り”の幾何平均を表す。

⁶ α 及び β は、日経225先物に対する超過収益率及び市場感応度を表す（回帰式「近似モデルの月次収益率 = $\alpha + \beta \times$ 日経225先物の月次収益率 + e 」により推定）。

表6 外挿データ期間における投資パフォーマンス

	収益率(年率)			α (年率)		β	
	AIトレーダー	近似モデル	日経225先物	AIトレーダー	近似モデル	AIトレーダー	近似モデル
外挿データ期間	19.19%	7.37%	14.21%	6.46%	-2.51%	0.8959	0.6953
2016年10月~2019年9月	13.39%	4.43%	11.51%	3.60%	-3.63%	0.8503	0.7005
2019年10月~2020年12月	32.33%	13.66%	20.69%	13.13%	-0.38%	0.9279	0.6788

図4 収益率の推移(外挿データ期間)



5 おわりに

本研究では、石原(2020)の研究で構築したAIトレーダーを“可読性の高い解釈可能なモデル”で表現し、その投資判断プロセスの説明を試みた。その結果、AIトレーダーの投資判断には「直近の日次収益率がプラスである場合には“売り”，マイナスである場合には“買い”を行う（リターン・リバーサル戦略）」、「直近のボラティリティが期待値より高い場合には“売り”，低い場合には“買い”を行う（高ボラティリティ局面でリスクを回避）」という特徴があることを読み解いた。また、作成したモデルの適合率と再現率の値から、その説明能力には改善の余地があるものの、AIトレーダーの投資判断をある程度表現できることを確認した。

参考文献

- [1] 石原龍太, “感度分析による AI トレーダーの投資判断アルゴリズムの解釈”, 先物オプションレポート 2020年1月号, (2020)
- [2] 笠井彰吾, “金融サービスをめぐる AI 活用の経緯 -AI 技術の現状を踏まえて-”, 立法と調査 2018.10 No. 405, (2018)
- [3] 加藤康之, “AI/フィンテックが変える資産運用”, フィナンシャル・レビュー 令和元年第4号, (2019)
- [4] 塩野剛志, “アセット・リターン予測 AI とマクロ経済理論の融合 -マルチタスク学習による正則化と識別-”, 第21回人工知能学会 金融情報学研究会 SIG-FIN-021-07, (2018)
- [5] 中窪文男, “金融資産価格のボラティリティーの推定と予測”, ニッセイ基礎研究所報 Vol. 22, (2002)
- [6] 原聡, “機械学習における解釈性”, 人工知能 33巻3号, (2018)

本資料に関する著作権は、株式会社大阪取引所にあります。

本資料の一部又は全部を無断で転用、複製することはできません。

本資料の内容は、株式会社大阪取引所の意見・見解を示すものではありません。

本資料は、デリバティブ商品の取引の勧誘を目的としたものではありません。