有明高専 情報システムコース
ゴーチェ研究室
内田陽太, 浦崎華瑠

「異常値を含むデータに対する文脈情報統合の有効性に関する研究」
― 株価予測を題材とした検証 ―

2025.10.22

目次

1．背景と課題

・背景

需給・市場・Web 履歴などあらゆるデータがログとして蓄積される現在、将来を見通す予測の必要性が一段と高まっている。

・課題

数値特徴だけでは、文脈依存的な「異常値」を説明することができない。

例：コロナショック、高市氏の総裁選出、
　　トランプショック

高市氏の総裁選選出時
日経平均終値（10月6日）
47,944.76
+2175.26

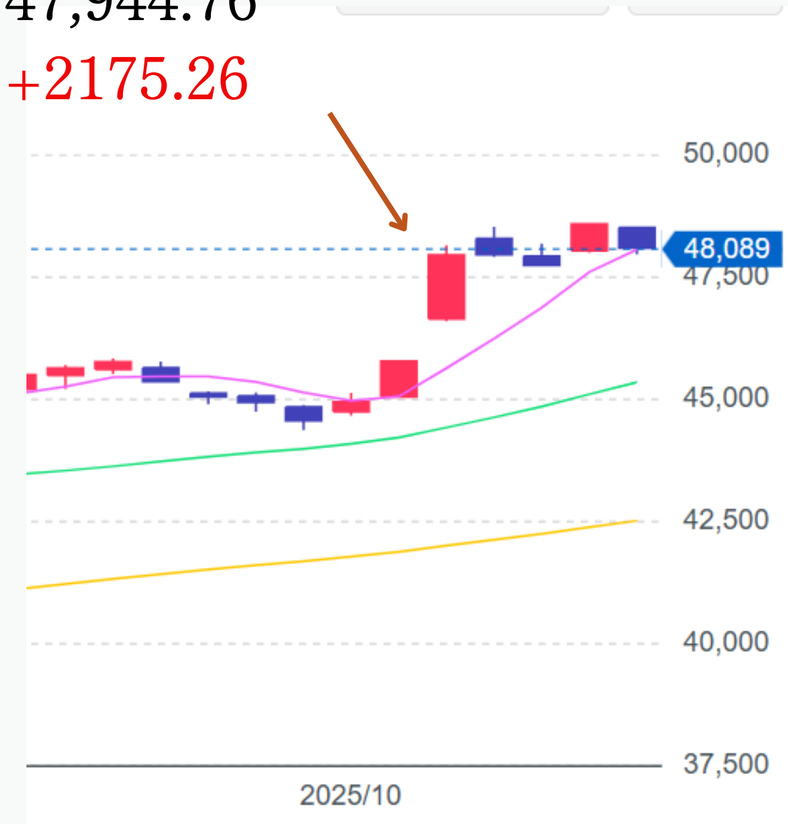


図1. イベントによる異常値の例

引用：日経平均株価：チャートーYahoo!ファイナンス
https://finance.yahoo.co.jp/quote/998407.O/char
(アクセス日10/14)

2．研究目的とプロセス

研究目的

- ニュース情報を統合した株価予測モデルの検討

  ニュースやイベントに含まれる文脈情報を数値データに統合し、異常値を「ノイズ」ではなく有用な情報として学習。
  性能を評価し、比較検証。

  大規模時系列データの代表例である「株価」を題材とする。

+

2．研究目的とプロセス

・研究の流れ

フェーズ1：数値特徴量による複数分類モデル構築

├ 線形（Logistic Regression）

├ 木構造（LightGBM）

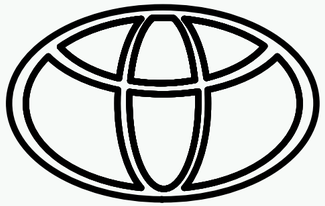
└ NN :ニューラルネットワーク（LSTM / Transformer）

上記の学習モデルを比較し、時系列データの予測に
有効なモデルを調べる。

フェーズ２．文脈情報の統合による汎化性能を評価

3．データの概要と特徴量

- **対象銘柄：** トヨタ自動車（7203.T）
- **期間：** 2001/01/01～2024/12/31
- **サンプリング単位：** 日足 (データ数 n = 5305)



- **予測対象**

1. main(全体)：テストデータ全体の集合
2. subset：　テストデータ全体から、翌日リターン比上位10%のデータ集合

- **特徴量**

価格・リターン系	ret1/5/10/20(1/5/10/20日リターン) devMA5/20/60(移動平均乖離)
市場指標系	topix_ret1, nikkei_ret1(各指数の1日リターン) vix_ret1(恐怖指数の日次変化)

etc.

4．目的変数と評価指標

- **目的変数の定義**

私たちが用いる目的変数は、株価の翌営業日の方向性を示した三値分類である。

以下にその目的変数を示す。

$$y_t = \begin{cases} 0 & (\text{下落}) \\ 1 & (\text{停滞}) \\ 2 & (\text{上昇}) \end{cases} \qquad (1)$$

4． 目的変数と評価指標

ラベル生成の手順を以下に示す。

予測対象は、1日先(H = 1)の累積リターンであり、次式で示される。

$$r_H(t) = \frac{Close_{t+H}}{Close_t} - 1 \qquad (2)$$

これを、以下に示す当日のボラティリティに応じてスケールした動的閾値 τ により分類する。

$$\tau_t(H) = k_\tau \times \sigma_{20}(t) \times \sqrt{H} \qquad (3)$$

$\left(\begin{array}{ll} \sigma_{20}(t) & : \text{当日までの20日間のリターン標準偏差} \\ k_\tau & : \text{ボラティリティに対する閾値スケール (定数)} \end{array} \right)$

4．目的変数と評価指標

(2)(3)式を利用し、私たちが使用した三値分類の定義を以下に示す。

$$y_t = \begin{cases} 0 & (r_H(t) < -\tau_t(H)) \\ 1 & (|r_H(t)| \leq \tau_t(H)) \\ 2 & (r_H(t) > \tau_t(H)) \end{cases} \qquad (4)$$

4．目的変数と評価指標

表１．具体例(翌日予測 H = 1)の場合

当日終値	翌日終値	r_H(t)	σ_20	τ_t(H)	クラス
2000	2010	+0.5%	1.5%	0.45%	2(上昇)
2000	1999	-0.05%	1.5%	0.45%	1(停滞)
2000	1980	-1.0%	1.5%	0.45%	0(下落)

4．目的変数と評価指標

- **評価指標**

多クラス分類ではクラス間の偏りを考慮する必要があり、「正解率」だけでは正しい評価はできない。

1.混同行列

表２．混同行列

予測\実際	下落(0)	停滞(1)	上昇(2)
下落(0)	TN	TN	FN
停滞(1)	TN	TN	FN
上昇(2)	FP	FP	TP

※上昇(2)をPositiveとしたとき

T: True

F: False

P: Positive

N: Negative

4．目的変数と評価指標

混同行列を用いて、計算される指標には以下のようなものがある。

precision(適合率) ⇒ 予測したうち正しかった割合
recall(再現率) ⇒ 実際に正しい中で当てられた割合

$$\mathrm{Precision}_i = \frac{TP_i}{TP_i + FP_i}, \quad \mathrm{Recall}_i = \frac{TP_i}{TP_i + FN_i} \qquad (5)(6)$$

各クラスごとの「適合率」「再現率」の調和平均を示す、F1スコアがある。

$$\mathrm{F1}_i = \frac{2 \times (\mathrm{Precision}_i \times \mathrm{Recall}_i)}{\mathrm{Precision}_i + \mathrm{Recall}_i} \qquad (7)$$

4．目的変数と評価指標

前述したF1スコアを多クラス分類に応用するために平均をとる。
F1スコアの平均をとったものをMacro-F1と呼ぶ。

$$\text{Macro-F1} = \frac{1}{3}\sum_{i=0}^{2}\text{F1}_i \tag{8}$$

また、Macro-F1以外に、補助的にBalanced Accuracyも用いる。
これは、recall(再現率)だけをみることで、公平性を確認している。

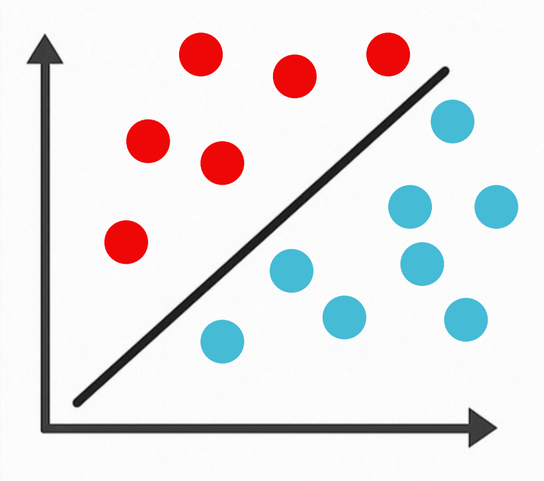
$$\text{Balanced Accuracy} = \frac{1}{3}\sum_{i=0}^{2}\frac{TP_i}{TP_i + FN_i} \tag{9}$$

5．各モデルの概要

・**非NN系**

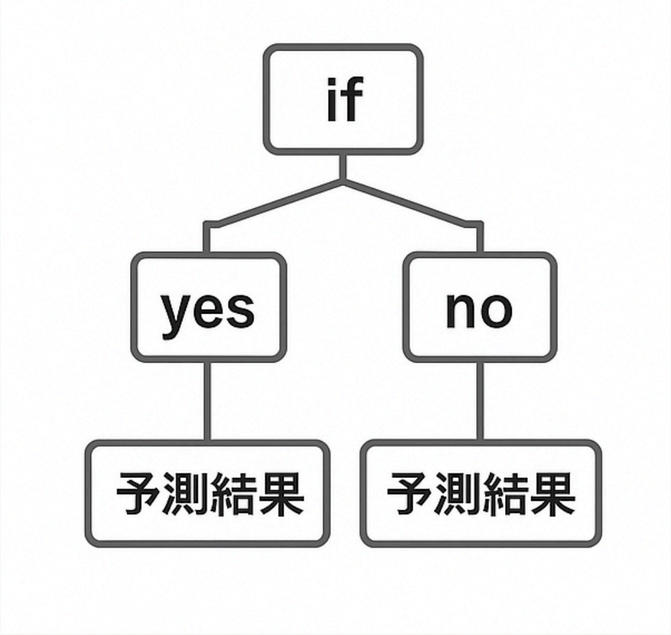
●線形系（Logistic Regression）

・入力と出力の関係が直線的なモデル
　y = ax + bのように表現される。
・非線形な関係は表現できない。



●木構造（LightGBM）

・決定木集合モデル
・非線形な関係を表現できる



5．各モデルの概要

・NN系（Deep Learning 系）

●Transformer

- 自己注意機構(Self-Attention)により
  系列内の重要な関係を学習
- 並列処理が可能（RNNは順次処理）
- 長期的な依存関係に強い

●LSTM

- RNNを改良したモデルで長期依存関係を保持
- ゲート構造を持つ
- 忘却ゲート：古い情報をどれだけ捨てるか
  入力ゲート：新しい情報をどれだけ入れるか
  出力ゲート：どの情報を次に渡すか

どちらも「時間の流れ」や「順序のあるデータ（文章・株価など）」
を入力として扱える。

5．各モデルの概要

・**Transformer, LSTMの入力形状**

入力: (B, T = 60, F = 36)
B: 各フォールドのサンプル数
T: 窓幅
F: 学習に使う特徴量数

出力: (B, 3)
各クラスごとの確率を出力する

表３．各フォールドごとの訓練/検証/テストデータ数

Fold	訓練データ	検証データ	テストデータ
1	785	151	1001
2	1634	363	1001
3	2483	575	1001
4	3332	787	1001

6．コード実装の概要

① データ

データの取得
及び整形

② 特徴量生成

特徴量を生成

③ ラベル付け

三値分類ラベル
を生成

④ 整形

log1p
inf/-inf,
NaNの処理

⑤ 分割

PWFE
によるデータ分割

⑥ 前処理

特徴量を絞る/
時系列を
窓へ変換

⑦ 学習

各モデルによる
学習

⑧ 評価

macro-F1
balanced
accuracy
混同行列

6．コード実装の概要（データ分割）

Purged Walk-Forward ＋ Embargoを採用することで、
「未来リーク」「自己相関バイアス」を防ぎ、汎化性能を現実的に評価する

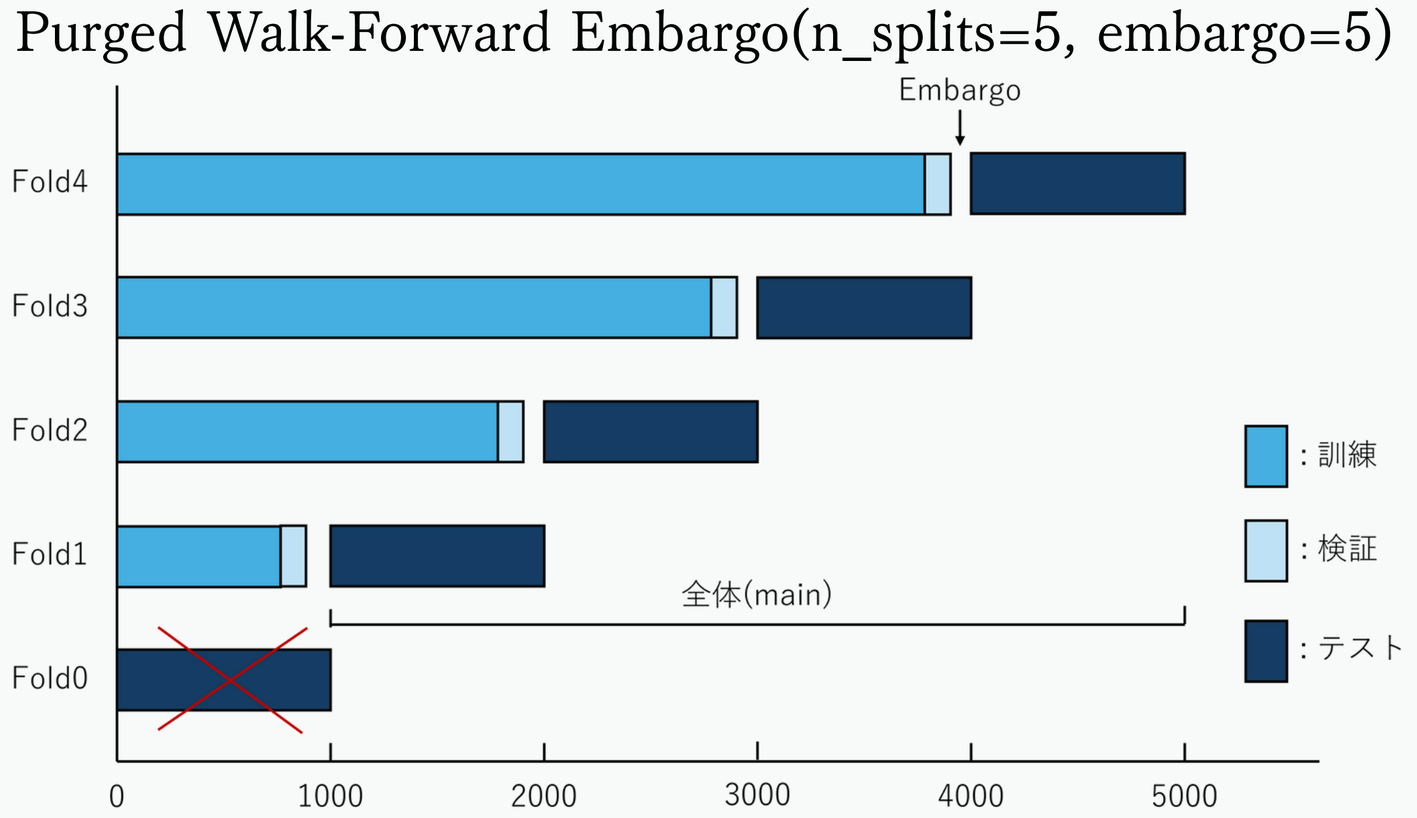


図2．データ分割

6．コード実装の概要（特徴量別重要度）

特徴量別重要度を可視化することで、どの特徴量が有用か、モデルを構成する際、取捨選択できる。

また、フェーズ2でニュースの文脈情報を特徴量として追加した際、どれほど精度に寄与したか可視化することができる。

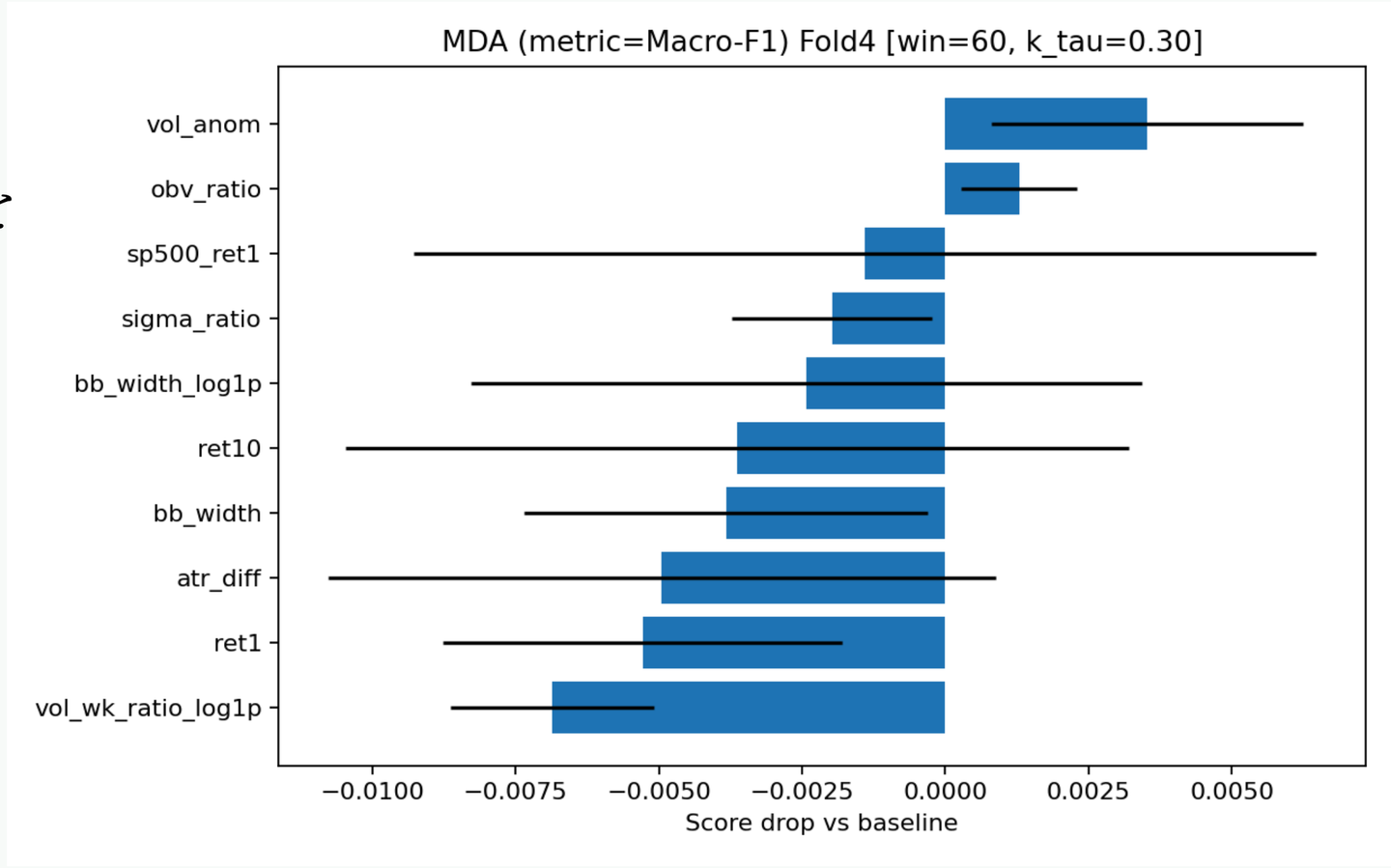
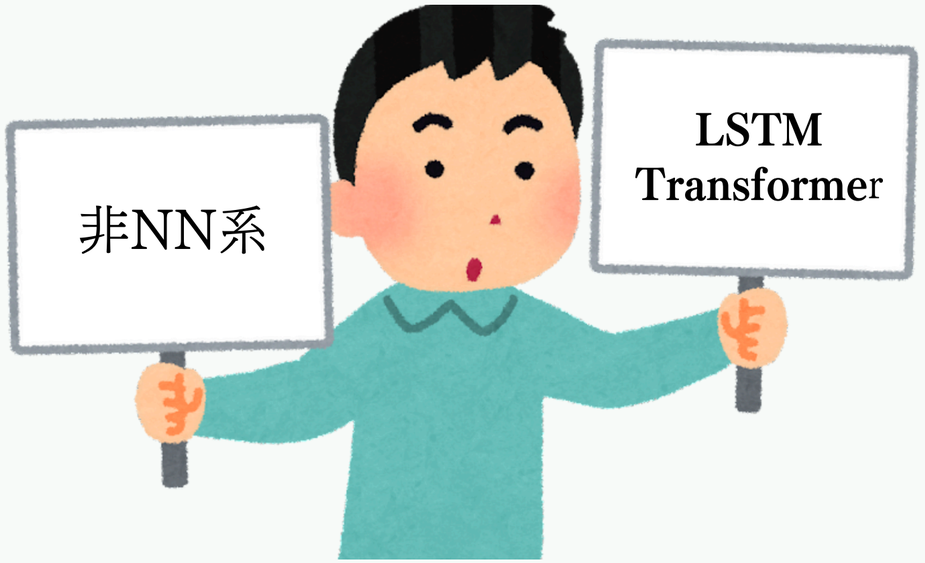


図3．transformerモデルのFold4における特徴量別重要度

7．予想

- **結果の予想**

時系列データに対する学習に強いLSTMやTransformerが非NN系よりも圧倒的に精度が良いものと予想。



8．研究結果

1．各モデルの実行結果

表4．実行結果（main）

モデル	macro_f1	balanced_acc
Transformer	0.3526	0.3561
LSTM	0.3485	0.3657
LightGBM	0.3534	0.3556
Logistic_Reg	0.3258	0.3284

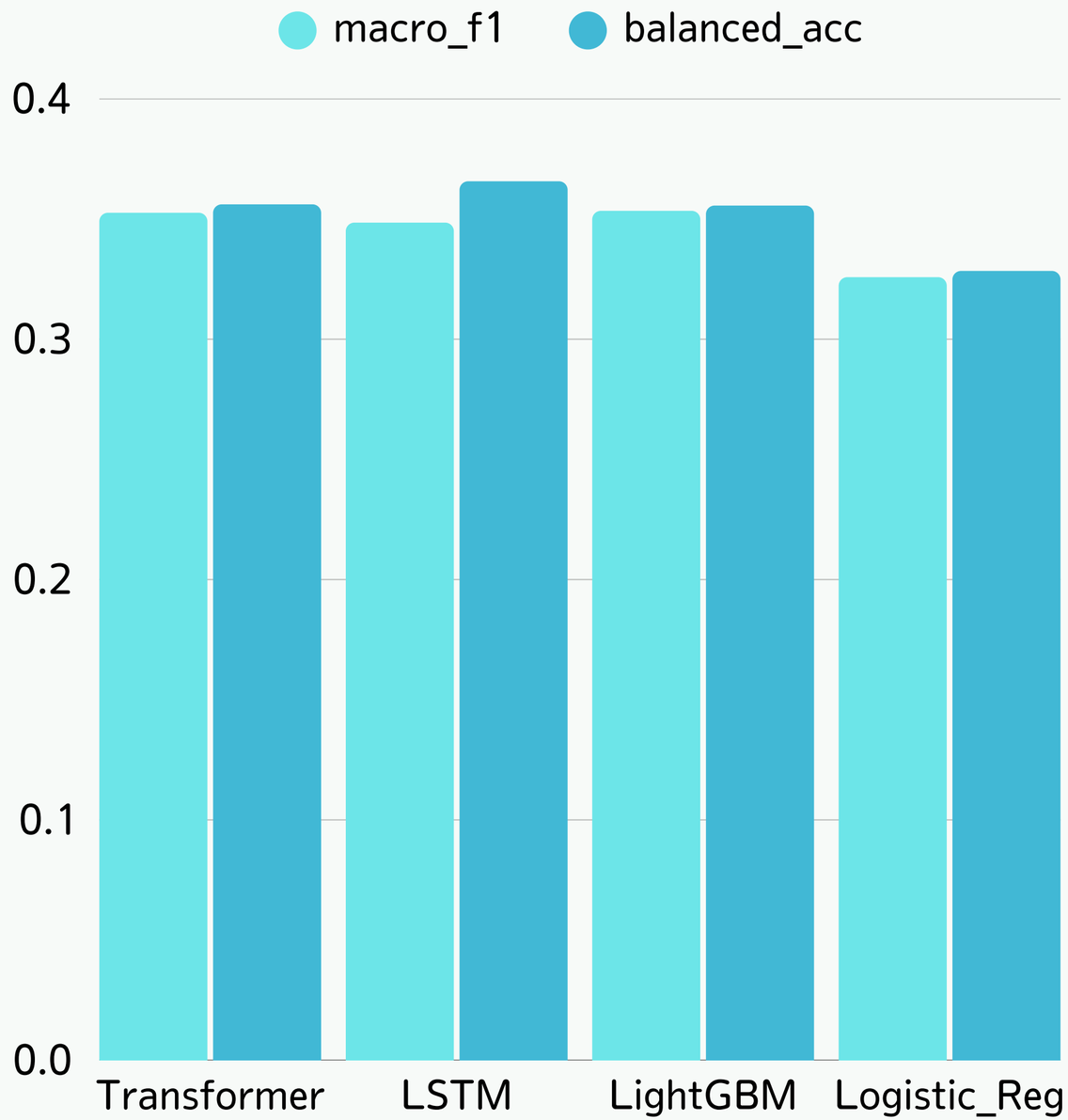


図4．実行結果（main）

8．研究結果

2．実行結果(subset)

表5．実行結果（subset）

モデル	macro_f1	macro_f1
Trasnformer	0.3037	0.4056
LSTM	0.2642	0.3474
LightGBM	0.2496	0.3200
Logistic_Reg	0.3106	0.4180

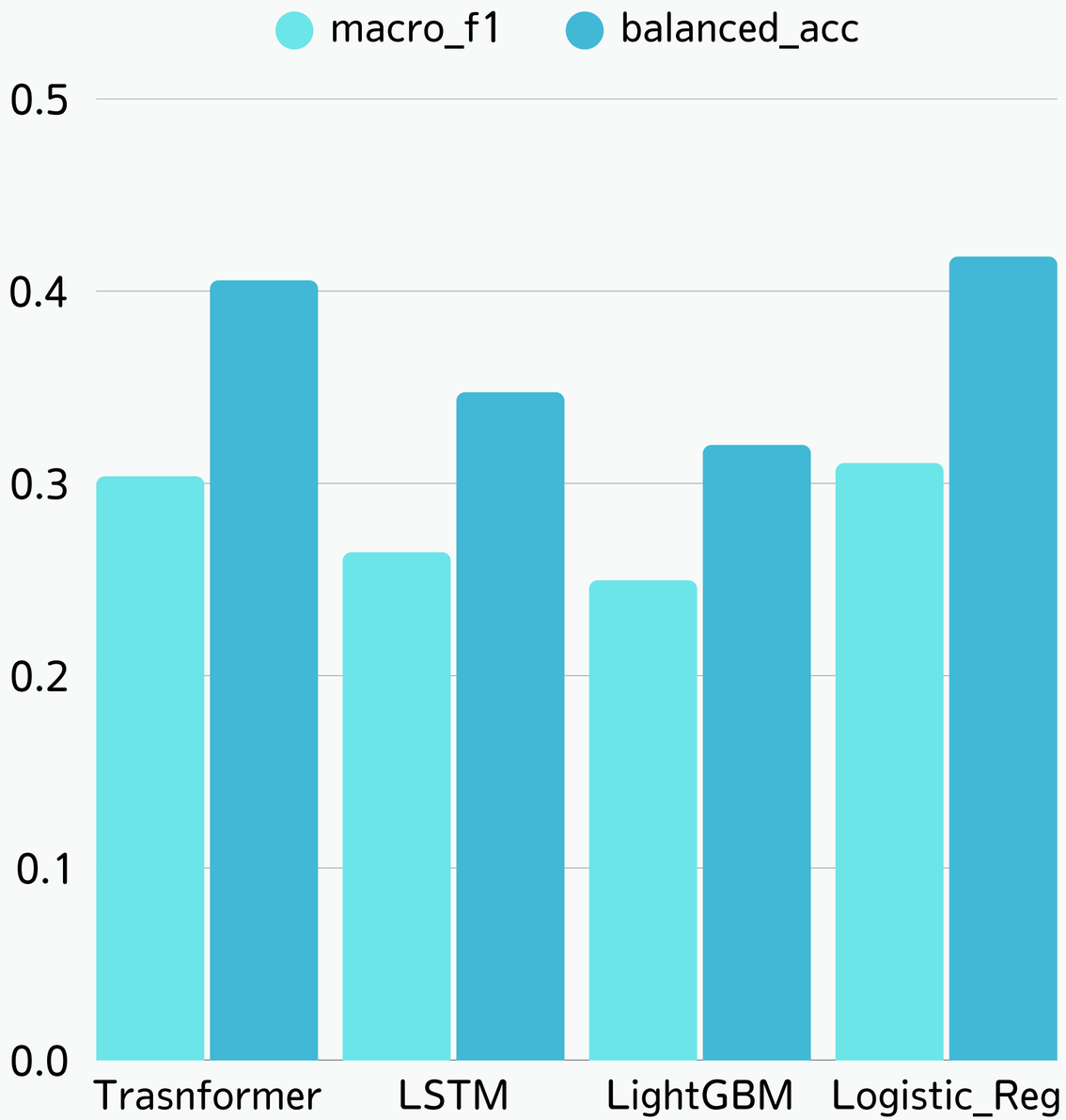


図5．実行結果(subset)

8．研究結果

3. 各フォールドの学習状況の可視化

学習曲線(train_lossとval_loss)およびval_accuracyをグラフとして出力

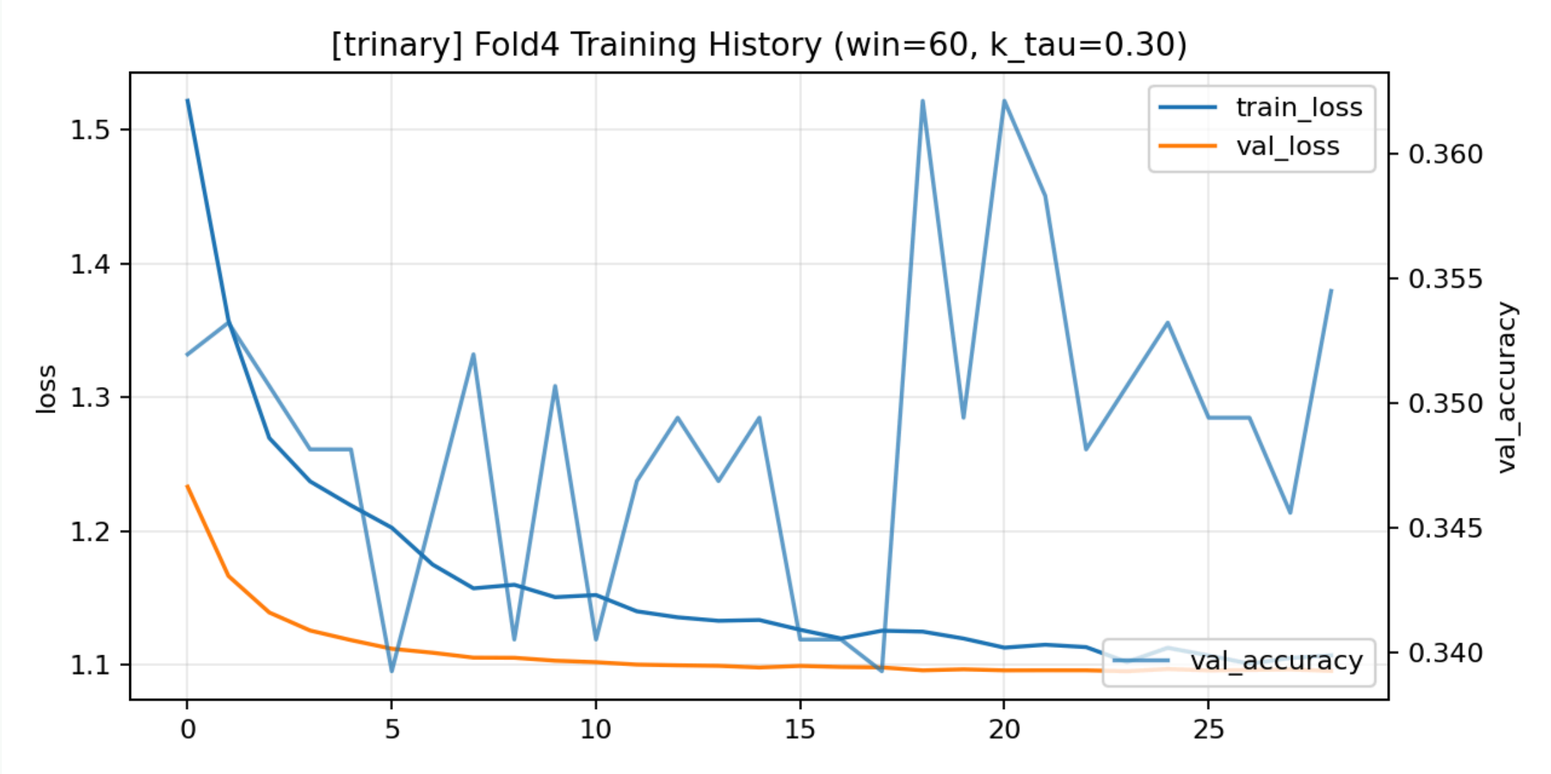


図6.　transformerモデルのFold4における学習状況

9．考察

○　なぜ時系列データに有効とされているNN系(LSTM, Transformer)が非NN系と横並びになっているか？

- 特徴量が価格派生中心ばかり
- ノイズの多さ
- サンプルデータ数が少ない

○　なぜ顕著にsubsetに対応できていないのか？

- 閾値の未調整：mainの設定を流用しているため、
  subsetの特性に最適化されていない
- 外部情報の不足：決算やイベントなど変動要因をとらえる特徴量が不足

9．考察

○　なぜsubsetで比較的TransformerとLogistic regressionが有効だったのか？

- LR：“方向手がかり”関連の特徴量の絶対値が大きくなりやすく、学習済みの直線境界に対するmargin（境界からの距離）が拡大。

- Transformer：self-attentionにより、直近の変化点や急騰区間へ重みが集中

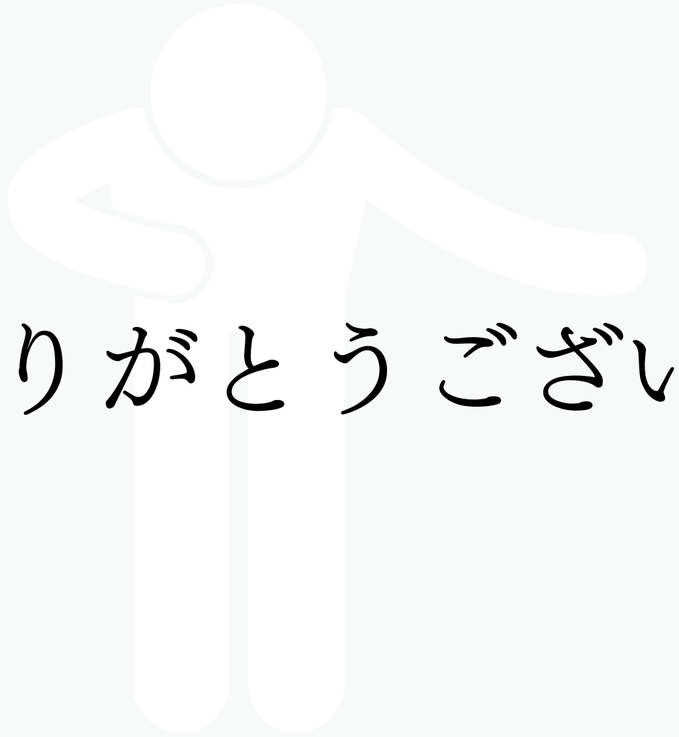
- LightGBM：“階段関数”だから強さが連続的に増えない

- LSTM：main向けのゲート調整が影響し、時系列依存を十分活かせなかった

１０．今後の展望

・NNの時系列データに対する予測精度が非NNと変わらない原因を探り、改善する。

・各モデルにニュースの「文脈情報」を特徴量として加えることで、汎化性能がどのくらい向上するか調べる(フェーズ2)



ご清聴ありがとうございました