

## 新規特徴量の提案一覧

以下では、既存の価格・テクニカル・需給・ファンダメンタル特徴量をもとに新たに生成可能な特徴量を、カテゴリ別に提案します。それぞれについて、派生元の既存特徴量、計算式（数式または擬似コード）、期待される効果、およびモデルや損失関数との相性をまとめました。重複や情報冗長を避け、既存特徴量と低相関で高い付加情報を持つものを優先しています<sup>1</sup>。各特徴量は実装しやすい形で記載しています。

### 組み合わせ特徴量（複数特徴の比・積・差など）

- **価格対移動平均比率**（Price/Moving Average Ratio）
  - 派生元: 終値 (Close), 移動平均 (例: 20日SMA)<sup>2</sup>
  - 計算式:  $\text{price\_to\_sma20} = \text{Close} / (\text{SMA20} + \epsilon)$ （終値を20日単純移動平均で割った値）<sup>3</sup>。他にも50日や100日移動平均との比率を検討可能。
  - 期待効果: 株価が中長期トレンドからどれだけ乖離しているかを定量化し、過熱感や反転余地を捉えます<sup>3</sup>。例えば値が大きければ短期的な加熱を示唆し、平均回帰的なリターンを期待できます。
  - モデル適性: 線形モデルでは直接過熱度を扱いやすく、非線形モデルでもトレンド局面の識別に寄与します（移動平均線との乖離はトレンドフェーズ検出に有効）<sup>3</sup>。ATFTでは時系列文脈中でトレンド成熟度を学習する助けとなり、中期的Sharpe向上に寄与すると期待されます。
- **RSI×ボラティリティ**（Momentum-Volatility Interaction）
  - 派生元: 相対力指数 (RSI14など), ボラティリティ (例: 20日標準偏差volatility\_20d)<sup>4</sup>
  - 計算式:  $\text{rsi\_vol\_interact} = \text{RSI\_14} * \text{volatility\_20d}$ （例えばRSIと年率換算ボラの積）
  - 期待効果: モメンタム指標であるRSIに価格変動率を掛け合わせ、**勢いの強さと変動の大きさ**を同時に考慮します。ボラティリティが高い中でRSIが高水準の場合、短期的な行き過ぎ（急騰・急落）シグナルとして捉えられ<sup>4</sup>、IC向上が期待できます。逆にボラ低下下の高RSIは緩やかなトレンド継続を示す可能性があります。
  - モデル適性: 非線形変換を施さずとも2変数の相互作用効果を直接提供でき、LightGBMのような木モデルでは分割条件に、この積特徴が効く可能性があります。ディープモデルにおいてもRSIとボラの関係性を明示することで学習を助け、極端局面での予測精度改善（リスクオフ/onの検知）に寄与します。
- **モメンタム×出来高**（Momentum with Volume Confirmation）
  - 派生元: 短期リターン (例: 5日リターン), 出来高系 (例: 出来高5日平均や出来高比率)<sup>5</sup>
  - 計算式:  $\text{vol\_confirmed\_mom} = \text{returns\_5d} * (\text{Volume} / \text{MA}(\text{Volume}, 20))$ （5日リターンと出来高20日平均比の積）など
  - 期待効果: モメンタムに出来高要素を組み込み、**出来高を伴った価格変動**を検出します。出来高が平均以上に増加している中での価格上昇はブレイクアウトの信頼度向上、一方出来高を伴う下落は需給悪化シグナルとなり得ます。単なる価格モメンタムより情報量が増し、ICや精度の向上が見込めます。
  - モデル適性: GBDT系では価格変化と出来高変化の相互作用を自動で捉えにくいため、明示的な積特徴は有効です。ディープラーニングでも、この特徴があることで「**出来高を伴った変動**」を直接捉えられ、ATFTモデルの短期予測やイベント時のロバスト性が向上すると期待されます。

## 変化率・加速度系特徴量（過去からの変化や勾配）

- **MACDヒストグラムの勾配** (MACD Histogram Slope)
  - 派生元: MACDシグナル・ヒストグラム (MACDh\_12\_26\_9 等) <sup>6</sup>
  - 計算式:  $\text{macd\_hist\_slope} = \text{MACDh\_today} - \text{MACDh\_yesterday}$  (MACDヒストグラムの1日差分。必要に応じて3日傾きなども検討)
  - 期待効果: モメンタム指標であるMACDの変化速度を表し、**トレンド転換点の早期検知**に有用です。ヒストグラムが縮小から拡大に転じる局面は上昇加速を、拡大から縮小に転じる局面は勢い減速を示唆します <sup>7</sup>。この勾配特徴により短期ICの改善やタイミング精度向上が期待できます。
  - モデル適性: ATFTのような時系列モデルでは勾配特徴自体を入力することで時間変化の学習を助け（勾配符号が正→負への変化を捉えやすい）、Turning Pointに敏感な損失関数（例えば方向性の分類誤差など）とも相性が良いです。GATにおいても各銘柄の勢い変化をノード特徴として与えることで、近隣伝播時に急変銘柄を強調できます。
- **出来高加速度** (Volume Acceleration)
  - 派生元: 出来高移動平均・比率 (例: volume\_ma5, volume\_ma20 等) <sup>5</sup>
  - 計算式:  $\text{volume\_accel} = (\text{Volume\_ma5} / \text{Volume\_ma20}) - 1$  あるいは  $\text{volume\_ratio\_5d} - \text{volume\_ratio\_20d}$  (直近出来高が中長期平均比でどれだけ増減したか)
  - 期待効果: 出来高の増加ペースを表し、**需給変化の加速**を捉えます。値がプラスなら出来高が急増中、マイナスなら減少中であり、急増する場合は市場の注目度上昇や新規フロー流入を示唆します <sup>7</sup>。これは転換点やブレイクアウトの前兆検知に繋がり、短期リターン予測に有益です。
  - モデル適性: 出来高水準とその変化を組み合わせるため、モデルは一次特徴だけでは捉えにくい**出来高トレンド**を利用できます。特に時系列モデルでは直近の出来高急増が将来の変動に与える影響を学習しやすくなり、GATでもセクター内での資金流入銘柄として当該ノードを際立たせる効果が期待できます。
- **ボラティリティ変化率** (Volatility Change)
  - 派生元: 短期・中期ボラティリティ (例: 5日、20日ボラティリティ) <sup>4</sup>
  - 計算式:  $\text{vol\_change} = \text{volatility\_5d} / (\text{volatility\_20d} + \epsilon) - 1$  (直近ボラが中期ボラからどれだけ変化したか。差分でも可)
  - 期待効果: **ボラティリティの急上昇・低下**を定量化します。急激なボラ上昇はリスクオフや材料出現を示唆し、今後の平均回帰的な動きやボラ収束を予兆します。一方ボラ低下局面からの跳ね上がりはブレイクアウトの兆候となり得ます。これらを特徴量化することで、中期Sharpe向上やリスクイベント前後の予測改善が期待できます。
  - モデル適性: 非線形モデルではボラの過去比較を自動学習できますが、明示することで学習効率が向上します。特にATFTでは過去のボラ推移と合わせて注意機構で重要視される可能性があり、リスク管理を組み込んだ損失（例えばボラティリティターゲット型の損失）ともシナジーがあります。

## クロスセクション特徴（同日時点での他銘柄との比較）

- **当日リターン順位** (Daily Return Rank)
  - 派生元: 当日リターン (returns\_1d) <sup>8</sup>
  - 計算式: 各営業日について、リターンを全銘柄内でランク化し0~1に正規化した値。例:  $\text{rank\_ret\_1d} = \text{rank\_percentile}(\text{returns\_1d by Date})$  (当日内の百分位) <sup>9</sup>
  - 期待効果: **横断面的な勝ち組・負け組の位置**を特徴づけれます。例えば当日トップ5%の急騰銘柄は翌日利益確定売りでリバース傾向、底辺5%の急落銘柄はテクニカルリバウンド狙いの買いが入る、と

いったアノマリーを捉えます。横断面での位置付け情報は個別銘柄の絶対値特徴では得られない追加予測力をもたらし、RankICの向上が期待されます。

- **モデル適性:** **Rank学習**との相性が良く、RankICを重視する目的関数では直接的な入力となります。Cross-Sectional Normalization（日次横断Zスコア）では市場全体効果を除去しますが<sup>10</sup><sup>11</sup>、本特徴量はモデルに「相対的な強弱」を明示的に渡すため、特に線形モデルやGATで各ノード比較を行う際に有益です。

- **出来高クロスセクションZ**（Cross-Sectional Volume Z-score）

- 派生元: 出来高（当日出来高、または出来高対平均比率）<sup>5</sup>
- 計算式:  $\text{vol\_cs\_z} = (\text{Volume} - \text{mean\_by}(\text{Date})[\text{Volume}]) / \text{std\_by}(\text{Date})[\text{Volume}]$ （当日全銘柄の出来高平均・標準偏差で標準化）<sup>11</sup>
- 期待効果: **当日に市場で突出して出来高が多いか少ないか**を測ります。プラスの大値なら「その日特に注目されている銘柄」、マイナス大なら「関心が薄い銘柄」を示唆します。注目度の高い銘柄はニュースやイベント発生の可能性が高く短期変動も大きくなりがちであり、一方極端に低い場合は材料難を意味します。この情報により情報係数の改善や予測分散の低減（無風銘柄の識別）が期待できます。
- モデル適性: 出来高自体はスケールの差が銘柄間で大きいいため、その日の中でのZスコアにより比較可能な指標にします<sup>10</sup>。モデルはこの特徴を用いて当日の市場注目度を織り込み、ATFTでは時系列的な出来高急変パターンと合わせて解釈し、GATではネットワーク内で異常に出来高が多いノードを強調して伝播させる効果が期待できます。

- **バリュエーション順位**（Valuation Percentile）

- 派生元: ファンダメンタル倍率（例: 予想PER, PBR等）
- 計算式: 例えばPERについて  $\text{pe\_percentile} = \text{rank\_percentile}(\text{PER by Date})$ （同業他社や全市場内での位置）
- 期待効果: **割高・割安度合いの相対比較**を提供します。同日において当該銘柄のPERが市場で上位に高ければ割高銘柄、下位なら割安銘柄であり、短期では割高銘柄ほど利益確定売り圧力を受けやすい等の効果が考えられます。横断面でのバリュエーション位置づけは中期リバージョン戦略（バリュエーション効果）に寄与し、中期Sharpe向上につながる可能性があります。
- モデル適性: ファンダメ情報は絶対値より相対値で解釈すべき場合が多く、特に**横断面学習**を行うモデルでは有効です。ライトGBM等でもこの順位特徴を分岐に使うことで非線形に割高/割安効果を捕捉でき、ディープモデルでもembeddingされたファンダメ指標と組み合わせて相対評価を学習する助けになります。

## 非線形変換・シグモイド特徴（閾値トリガー型の変換）

- **過熱RSIシグモイド**（Sigmoid of Overbought RSI）
- 派生元: RSI14などモメンタム指標<sup>6</sup>
- 計算式:  $\text{rsi\_overheat\_sig} = \text{sigmoid}((\text{RSI\_14} - 70)/5)$ （RSIが70を超える領域で0.5を超えて急激に1に近づくシグモイド）
- 期待効果: **モメンタム指標の閾値効果を滑らかにエンコード**します。RSIが70を超えると一般的に過熱圏と判断されますが、シグモイド変換により「70付近で急激に値が上昇する特徴量」を作ることで、モデルが非線形な臨界現象を捉えやすくなります。これにより過熱時の反転リスク検知が強化され、情報係数の改善やドローダウン低減に繋がる可能性があります。
- モデル適性: 人工的に非線形閾値効果を前処理で埋め込む手法であり、線形モデルでも閾値効果を取り込めます。ディープモデルでは本来ネットワーク内部で学習できる非線形ですが、明示的特徴として

与えることで学習を安定化できます。特に**極端値に対する損失関数**（例えば異常検知的なペナルティ）を用いる場合に相性が良いです。

#### ・ボラ急騰フラグ (Volatility Spike Flag)

- 派生元: 短期ボラティリティ指標（例: 20日ボラ）、またはその日の日次リターン <sup>12</sup>
- 計算式: `vol_spike_flag = (volatility_20d > vol_threshold).int8()` あるいは日次リターン絶対値が平常時の2σを超えたかなど <sup>12</sup>
- 期待効果: **急激な変動発生を閾値で捉える**二値特徴です。例えば直近ボラが過去半年平均の2倍を超えた場合に1とすることで、平常と異なるボラティリティレジームに入ったことを示します <sup>12</sup>。これはモデルに対し「平常時と異なる振る舞いをすべき局面」であることを教える働きをし、そうした局面での予測精度向上やリスク管理に資する行動（ポジション縮小等）が期待できます。
- モデル適性: 閾値をまたいだ非線形な挙動を明示的に入力するため、ツリー系モデルでもルールベースのように捉えられます。ATFTモデルではボラ急騰フラグを時間的マスクや重み調整に活用でき、マルチホライズン損失でも急変時の予測を別扱いするような重み付けと組み合わせると効果的です。

#### ・ゴールデンクロス/デッドクロス指標 (Moving Average Crossover Flag)

- 派生元: 長短期移動平均（例: 50日線, 200日線） <sup>2</sup>
- 計算式: `golden_cross_flag = (SMA50 > SMA200).int8()`（50日平均が200日平均を上回ったら1、下回れば0。デッドクロスは逆）
- 期待効果: **長期トレンド転換のシグナル**を捉えます。ゴールデンクロス発生時は上昇トレンドへの転換点として注目され、一定期間ポジティブなモメンタムが続く傾向があります。一方デッドクロスは下降転換シグナルです。これらを特徴量化することで、中期的なトレンドフォロー/逆張り判断に寄与し、中期SharpeやHit率の向上が期待できます。
- モデル適性: 非線形かつヒステリシスのあるパターンを単一のフラグで表現できます。ディープラーニングモデル内部で畳み込みや注意機構により検出可能ですが、明示することで**学習データが少ない場合でもルールベース知識**を提供できます。損失関数が短期・中期の複合であれば、中期予測部分の精度改善に効くでしょう。

## 時系列・カレンダー要素特徴

#### ・曜日ダミー (Day-of-Week One-Hot)

- 派生元: 日付（曜日情報） <sup>13</sup>
- 計算式: 週の各曜日（月～金）に対しOne-Hotエンコード。例: `is_mon, is_tue, ... is_fri` としてそれぞれ該当曜日なら1 <sup>13</sup>。
- 期待効果: **曜日アノマリー**を捉えます。月曜は週末の悪材料出尽くしで下落しやすい、金曜は持ち越し回避で上昇しづらい等、統計的に弱い傾向が指摘されています <sup>14</sup>。モデルに曜日情報を与えることで微小な平均リターン差を取り込み、特に1日先予測のICや精度を僅かに押し上げる効果が期待できます <sup>15</sup>。
- モデル適性: カレンダー要因は他特徴と独立で、線形モデルでも容易に扱えます。ATFTのような時系列モデルでは季節性を学習するのに役立ちますし、複数期間の損失を用いる際には**曜日ごとのリスクプレミアム**を予め補正する役割も果たします（例: 月曜予測のバイアス除去）。

#### ・月中・月末フラグ (Month-End Timing Flags)

- 派生元: 日付（カレンダー位置）
- 計算式: 月初・月中・月末を区分するカテゴリー特徴。例: 月末5営業日以内なら `month_end_flag=1`、月初5日なら `month_start_flag=1`、その他0など。

- 期待効果: **月替わりアノマリー**や**リバランス効果**を捉えます。月末は機関投資家のリバランスやドレッシング買いで価格が変動しやすく、月初は資金流入で株高になりやすいという経験則があります。これらの時期をモデルに意識させることで、特定週における予測エラーの低減やSharpe改善が期待できます。
- モデル適性: 時系列モデルでは内部でこの周期性を学習可能ですが、明示することでデータ期間が短くても対応できます。特に**決算発表集中期**（日本では3月/4月や11月など）は月末月初と重なることも多く、そうした複合イベントのハンドリングで損失関数（例えば期間ごとに異なる損失重み）と組み合わせると効果的です。
- **決算発表前後距離** (Days to/from Earnings)
  - 派生元: 決算カレンダー（日付情報）、株価
  - 計算式: 次回決算発表予定日までの日数、および直近決算からの経過日数をそれぞれ数値特徴量化。  
例:  $\text{days\_to\_earnings} = (\text{earnings\_date} - \text{Date})$ （営業日ベース）
  - 期待効果: **決算イベントによる影響**をモデルに織り込めます。発表前は思惑買い/ポジション解消で価格変動が大きくなりやすく、発表直後はサプライズに応じたドリフト（Post-Earnings Announcement Drift）が数日続く傾向があります。この特徴によりイベントドリブな価格変動を説明でき、予測残差の低減やイベント時のSharpe向上が期待できます。
  - モデル適性: 決算発表日は未来情報にならない範囲で事前に公表されているためリークではありません。時系列モデルではこの特徴をキーに発表前後のパターンを異なる状態として処理できます。特にマルチホライズン予測で、1日先と5日先で決算を挟むか否かで難易度が変わる場合などに、Loss関数で重み調整する際の条件にも使えます。
- **SQ週フラグ** (Derivatives Settlement Week)
  - 派生元: 金融カレンダー（毎月第2金曜＝SQ日など）
  - 計算式: 各月の先物・オプション清算日（SQ）を含む週に1を立てる。日本株の場合、原則毎月第2金曜日がメジャーSQ（特に3,6,9,12月は四半期SQ）。例: その週の金曜がSQなら当該週の  $\text{is\_SQ\_week}=1$ 。
  - 期待効果: **先物・オプション清算に伴う特殊需給**を反映します。SQ直前は需給主導の売買（ヘッジポジション解消やロールオーバー）が発生しやすく、現物株にも影響が出ます。特にボラティリティや出来高が増加しやすいため、このフラグでモデルに警戒させることで、予測分布の不確実性を適切に評価したり、大きな外れ予測を防ぐ効果が期待できます。
  - モデル適性: GATなど銘柄間関係を扱うモデルでは、一斉に市場構造が変化するSQ週情報は全ノード共通のグローバル情報として有効です。またリスク管理を組み込む損失関数（例えばSQ週はペナルティを強める等）と組み合わせ、モデルがこの週だけ異常値に対応できるよう誘導することも可能です。

## セクター情報との結合特徴

- **セクター平均乖離** (Deviation from Sector Mean)
  - 派生元: 任意の数値特徴量 (例: 5日リターン, PER, ボラティリティなど) とセクター集約値 <sup>16</sup> <sup>17</sup>
  - 計算式:  $\text{feat\_vs\_sector} = x_i(t) - \text{mean}_{\{j \in \text{sector}(i)\}}(x_j(t))$ （当日同一セクター銘柄の平均からの差分）や、標準偏差で割ったZスコア版も考慮 <sup>18</sup>。
  - 期待効果: **同業他社と比べて突出している度合い**を定量化します。例えば5日リターンのセクター平均乖離は、「セクター全体は横ばいだがこの銘柄だけ大きく上昇している」といった個別要因の存在を示唆します <sup>17</sup>。これにより共通要因を差し引いた固有の強さ弱さを捉えられ、短期予測ノイズの低減やIC向上が期待できます。PERや利益成長率などファンダ指標でも同業平均との差をとることでバリュエーション面での乖離を特徴づけられます。

- モデル適性: セクター平均との差分は**共通因子を除去**する役割があります<sup>17</sup>。モデルはこの特徴を利用して市場全体やセクター全体の動きに対する相対的なパフォーマンスを学習でき、特にGATではセクター内の情報伝搬時に平均乖離が大きいノードを異常値として扱うなどの効果も期待できます。ATFTでも入力特徴量から共通成分を減じているため学習が安定しやすくなります。

#### • セクター内順位 (Rank within Sector)

- 派生元: リターンや成長率等の各種指標+セクター情報<sup>17</sup>
- 計算式:  $\text{rank\_in\_sector} = \text{rank\_percentile}(x\_i(t) \mid \text{sector}(i), \text{Date}=t)$  (当日同セクター内での百分位ランク)
- 期待効果: **セクター内での相対的ポジション**を示します。同じ業種内でトップクラスに値上がりしている銘柄なのか、あるいはアンダーパフォームしているのかを特徴づけることで、セクターローテーションやペアトレード的なリバージョン効果をモデルに捉えさせることができます。例えばセクター内で突出して上昇した銘柄は利食い売りにより短期的に平均回帰しやすい、といった傾向を捉え、RankICの改善やポートフォリオの分散効果向上に寄与します。
- モデル適性: 横断面順位をセクター毎に細分化することで**細粒度な相対比較**が可能になります。GATでは各セクター内サブグラフでのノード特徴として順位を持たせると、情報共有の際にリーダー銘柄・出遅れ銘柄を区別しやすくなります。ATFTでもセクター内での立ち位置を明示することで、異なるセクター間での学習を統合する際のばらつきを減らし、損失関数のセクター偏り（特定セクターだけ予測が難しい等）を是正する助けとなります。

#### • セクター内出来高シェア (Volume Share in Sector)

- 派生元: 出来高, セクター情報<sup>5</sup>
- 計算式:  $\text{volume\_share\_sec} = \text{Volume} / \sum_{j \in \text{sector}(i)} \text{Volume}_j$  (当日同セクター内出来高合計に占める割合)
- 期待効果: **その銘柄がセクター内でどれだけ注目されているか**を示します。出来高シェアが平常時より大きければ、「同業他社に比べこの銘柄に資金が集中している」状態であり、個別材料やマーケットテーマの中心にいる可能性があります。高い場合は短期的なモメンタム継続や過熱のシグナル、極端に低い場合は物色圏外でトレンド不参加のシグナルとなり得ます。これにより短期のアルファシグナルの精度向上が期待できます。
- モデル適性: セクターというグループ内での相対的な資金フローを直接数値化した特徴であり、ディープモデルの学習にも有益です。GATではノードの重み付けとして機能し、セクター内で重要度の高いノードを強調する効果があります。マルチタスク学習や損失調整の場面でも、セクター内シェアが低い銘柄（出来高希薄で予測困難）は損失寄与を下げるなどの応用が考えられます。

## GAT・時系列変換と相性の良い構造化特徴

- **グラフ次数 (関連ネットワーク次数)** (Graph Degree Centrality)
- 派生元: 銘柄間関連ネットワーク (例: 相関行列の隣接行列)<sup>19</sup>
- 計算式: ノードに接続するエッジ数または重み合計。例: 相関係数0.3以上の他銘柄数をカウントした  $\text{corr\_degree}_i = \text{count}_j([\text{corr}(i, j) > 0.3])$  や、上位kエッジ数でも可。
- 期待効果: **株式ネットワークにおける中心性**を表します。次数が高い銘柄は他と値動きが類似しやすい大型株・インデックス寄り度の高い銘柄である可能性が高く、ネットワーク上のハブとして振る舞います。一方、次数が低い銘柄は独立したテーマ株・材料株であり、相場全体とは異なる動きをする傾向があります。この特徴により、モデルは銘柄ごとの構造的な役割を認識し、GAT部分の**伝搬強度**を調節できるようになります。
- モデル適性: GATでは各ノードの初期特徴として次数を与えることで、注意機構が「このノードは繋がりが多い/少ない」という前提知識を持てます。次数の高いノードはメッセージパッシングを強く受け

る/与えるような学習が進み、結果的にグラフ伝搬の安定性と精度向上が期待できます。ATFTには直接関係しませんが、銘柄固有の構造情報としてembeddingに組み込むことで、時間変化と相互作用する複合パターンを捉えやすくなります。

#### • 近傍平均相関 (Average Peer Correlation)

- 派生元: 相関行列 (直近60日など), ピア (近傍) 銘柄リスト <sup>20</sup> <sup>21</sup>
- 計算式: 自銘柄とk近傍銘柄との相関係数の平均。例: `peer_corr_mean = mean_j(corr(i, j) for j in top_k_corr_peers(i))`
- 期待効果: **その銘柄がどれほど周囲と同調して動いているか**を示します。値が高ければ「市場全体・セクターと一蓮托生の銘柄」(ベータが高くアルファが出にくい)、低ければ「独立した値動きをする銘柄」(ユニークなアルファが期待できる)と解釈できます。この特徴は、モデルに対して予測可能性の事前情報を与える効果があり(独立株は予測困難だが連動株は指標などで予測しやすい)、マルチホライズンSharpeの安定性向上やモデルの不確実性見積もり精度向上に寄与します。
- モデル適性: GATでは隣接ノードからの情報集約時に、この平均相関を用いて**重み付け調整**を行うなどの工夫が可能です(相関が低いノードの情報は飛ばし、高い場合はより強く伝える等)。特徴量として持たせても各層で学習可能ですが、明示することで初期段階から「このノードは周囲と同期的/非同期的」というヒントを与えられます。ATFTにおいても銘柄固有の動きの強さを表す特徴として、時間方向の注意機構と組み合わせり予測に貢献します。

#### • 連騰・連敗日数 (Streak Length)

- 派生元: 価格連騰・連敗情報 (終値データ)
- 計算式: 直近連続で上昇した日数、下落した日数を計算。例: `win_streak = 連続陽線日数`, `lose_streak = 連続陰線日数 (当日まで)`
- 期待効果: **モメンタム継続性と反転ポイント**を特徴づけます。連騰記録が長いほど短期的な過熱感や利食い圧力が高まり、逆に連敗が続けば自律反発の期待が高まるという経験則があります。この特徴はまさにそれを数値化し、短期のリバースيون傾向をモデルに捉えさせます。結果として1日先の方向予測精度向上や極端局面でのSharpe改善が期待できます。
- モデル適性: 時系列モデルでは内部状態で連騰・連敗を把握できますが、明示的特徴があることで**より直接的にその効果を学習**できます。特に短期リターンに重みを置く損失関数では、この特徴を使って急騰・急落後の反転を当てにいく動きを強めることができます。GATにおいても、連騰中の銘柄ノードは近隣に波及効果を与える可能性があり、そのような情報を各ノード特徴に含めることは有益です。

---

各提案特徴量は、既存の特徴量セットと重複しない新たな情報を提供することを重視しています。例えば**クロスセクション順位**や**セクター平均乖離**は既存の単独特徴を相対化して情報独立性を高めるものです <sup>10</sup> <sup>17</sup>。また**非線形トリガー型**や**構造化特徴**はモデルが自力では抽出しにくいパターンを事前に供給し、学習効率と予測力を高めます。以上の特徴量群を組み込むことで、情報係数の底上げやSharpe向上が期待でき、ATFT+GATモデルやマルチホライズン損失とのシナジーも得られるでしょう。各特徴量はデータ漏洩に注意しつつ(例: 当日終了時点までの情報のみ使用)、実装可能ですので、優先度や効果検証を踏まえて順次適用をご検討ください。 <sup>10</sup> <sup>7</sup>

---

1 10 **glossary.md**

<https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/docs/glossary.md>

2 3 4 5 6 8 11 12 16 17 18 **dataset\_new.md**

[https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/docs/ml/dataset\\_new.md](https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/docs/ml/dataset_new.md)

7 **FEATURES\_margin\_weekly.md**

[https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/docs/FEATURES\\_margin\\_weekly.md](https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/docs/FEATURES_margin_weekly.md)

9 19 20 21 **peer\_features.py**

[https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/src/features/peer\\_features.py](https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/src/features/peer_features.py)

13 14 15 **DEEP\_RESEARCH.md**

[https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/docs/DEEP\\_RESEARCH.md](https://github.com/wer-inc/gogooku3/blob/80c6c2e43c6b7f919a1eee277e1b0516f9436760/docs/DEEP_RESEARCH.md)