

## 吉野貴晶 のクオンツ トピックス : NO11 画像認識 (CNN) を用いたモメンタム効果のパス情報の取り込み

### AI/機械学習によるモメンタム効果の改良

- 連載形式でAI (人工知能) と投資手法の関係性を紹介。
- 引き続き、クオンツ手法の一つであるモメンタム/リバーサルが今回のテーマ。

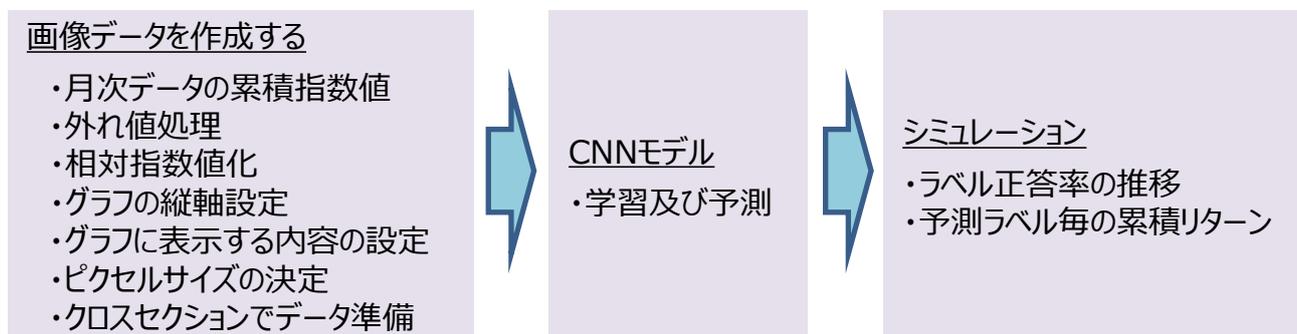
最近、AI (人工知能、以下AI ) に関連するニュースが増えています。投資の分野でも研究開発が盛んに行われており、実際に投資手法として利用可能な段階まで進展しています。本レポートでは、AIと投資手法の関係性をご紹介したいと思います。

前回レポートから引き続き、「クオンツとAI/機械学習の融合」をテーマにします。

投資の世界には従来からクオンツという職種が存在します。具体的には、統計解析等の数学的手法を用いて、マーケットの分析や投資戦略の構築をするのが主な仕事になります。No9のレポートでは、クオンツ領域において有名な投資手法であるモメンタム/リバーサル効果の概論をご紹介しました。No10のレポートでは、モメンタム/リバーサル効果の強化 (エンハスト) 案として、最もシンプルな機械学習モデルである決定木モデルを取り上げ、中期モメンタムを入力データとしてパフォーマンスのシミュレーションを行いました。その際、入力データの加工方法による結果の差異も併せて検証しました。

今回も引き続き、モメンタム効果を機械学習によってエンハストすることをテーマとします。今回は、現在の機械学習の主流である画像認識技術 (CNN、畳込みニューラルネットワーク) を取り扱います。また、データとしては、数値情報である個別銘柄のリターン値から画像を作成していきますが、その際に幾つかのデータ前処理を挟みます。最終的には、予測分位毎のリターンシミュレーションまで掲載します。

図1. 今回のレポートの流れ



## モメンタムのパス（経路）は有効な情報か？

### 1. モメンタム効果におけるパス（経路）情報の取り込み

従来からクオンツ領域で良く扱われる手法の一つとして、モメンタム効果があります。これは、相対的に株価が上昇（下落）傾向にある銘柄は、将来も上昇（下落）を続ける、という順相関の関係性に着目した投資戦略です。計算方法は以下の図2にあるように、過去のリターンそのものに着目した、ある意味シンプルな戦略となっています。シンプルですが、過去世界中の市場や資産クラスでその効果の存在が確認されており、広く出現するアノミー、と言えます。本レポートでは中期モメンタム（12ヶ月リターン、1か月スキップ）をモメンタムとして扱います。純粋にリターン値を扱っているため、この測定期間（12ヶ月リターン、1か月スキップ）において、株価のパス（経路）がどのようなものであったかは考慮されていません。図3に、同じ上昇幅だが、そこに至るまでの経路が違う場合のイメージを載せています。このように、始点と終点の株価差のみに着目する場合、株価のパス（経路）は考慮されません。今回は、このパス情報を取り込むことを目指します。

図2. 12か月（中期）モメンタム

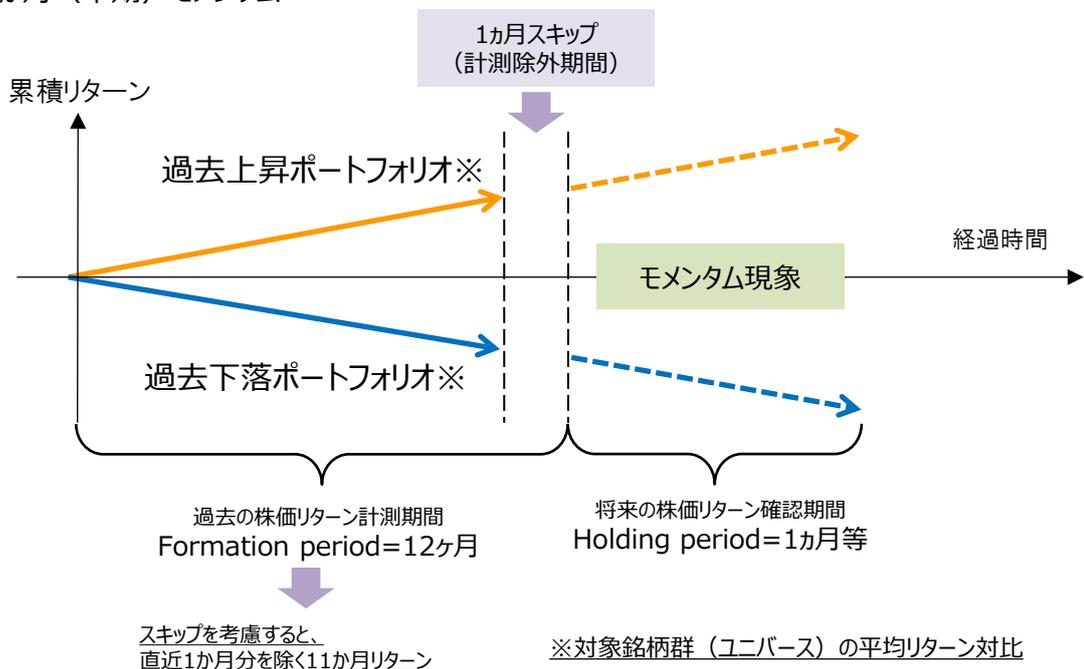
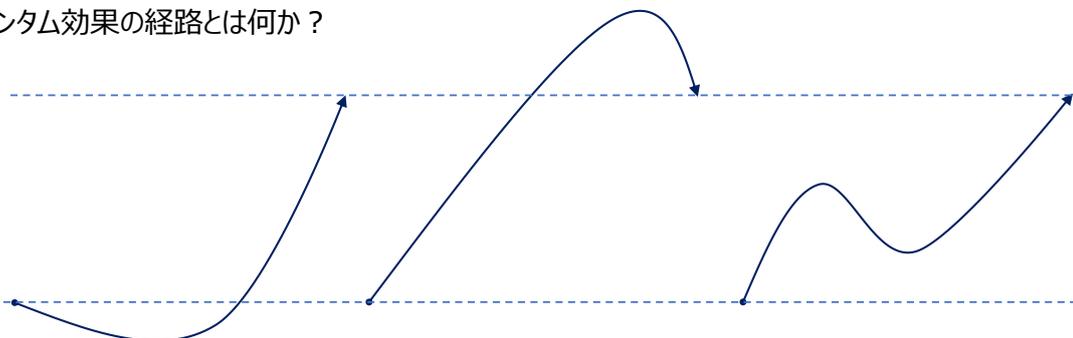


図3. モメンタム効果の経路とは何か？



- ✓ 同じ期間において、最終到達点（リターン）は同じ
- ✓ 一方、最終到達点に至るまでの途中の経路は3つとも大きく異なる
- ✓ 果たして、この途中経過の情報はリターン予測において意味があるのか？

## 画像データを作成する

### 2.データ作成

画像認識に取り組む際の最初の関門として、データそのものが準備出来るか？という問題があります。今ではインターネット上で様々な画像データが機械学習用として提供されていますが、今回作成したい画像、言わばモメンタム効果用の画像は手に入りません。そこで、自分で画像を準備する必要があります。

#### 手順1：データの抽出

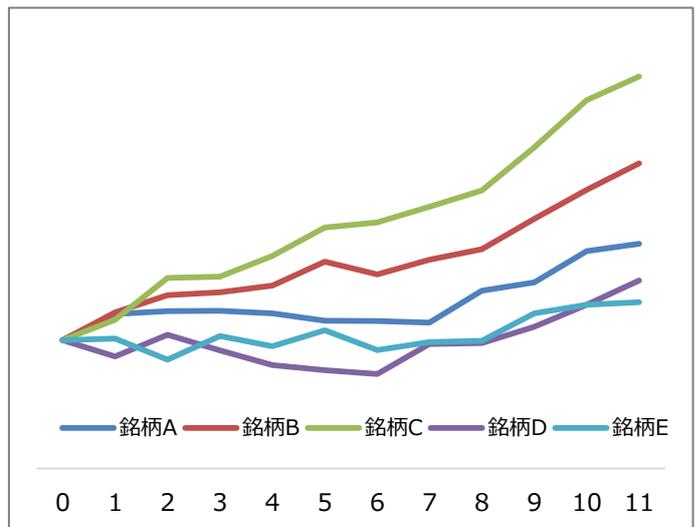
弊社内のデータベースを利用し、各銘柄の月次リターン値を作成します。なお、月次データを今回用いる理由は、日次よりもノイズが少ないと仮定したためです。

#### 手順2：累積指数値の作成

銘柄毎に13か月分のリターンを抽出します。始点を100として、累積指数値を作成しますが、この際、中期モメンタムの計算概念に則り、11か月までの累積リターンを作成します。

12ヶ月目のリターンは1か月スキップに該当するので利用しません。この12ヶ月リターン1か月スキップによる中期モメンタム効果に対応する将来のリターンは、13か月目の月次リターンとします。

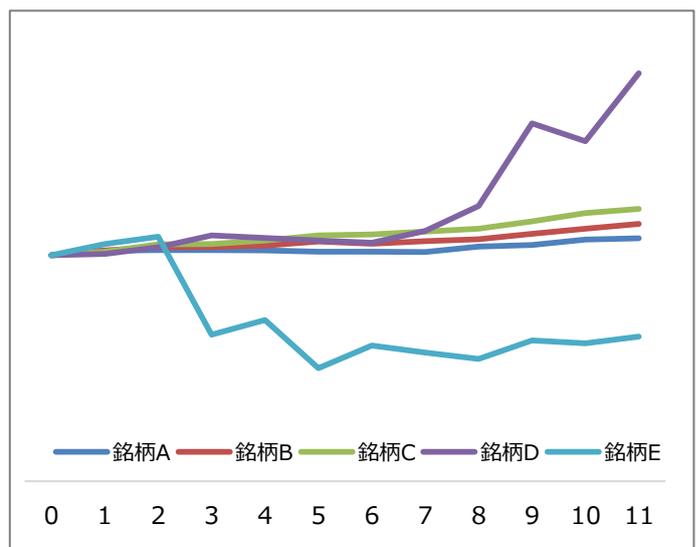
図4.過去12ヶ月前を始点とした11ヶ月累積リターン指数値



#### 手順3：外れ値処理

最終的には手順2で作成した指数値をグラフ化したものを入力データとしますが、先に外れ値処理を実施します。理由としては、グラフ化した際の折れ線の潰れを無くしたいためです。図5にあるように、もし相対感を持たせてグラフ化した場合に、極端に大幅な上昇、下落の銘柄があると、ほかの大多数の銘柄の折れ線グラフが潰れてしまう可能性があります。この状況を考慮するために、該当期間（11か月までの累積指数値）における全銘柄の指数値の最大値、最小値を抽出します。最大値を銘柄間で横比較した際に、上位2.5%よりも大きい銘柄は今回はデータから除外します。同様に、最小値が下位2.5%以下の場合も除外します。極端な動きをした銘柄は両基準に引っかかる場合がありますが、重複が無いとすると、全体の5%が除外されます。

図5.大幅上昇 or 下落銘柄があった場合のグラフ

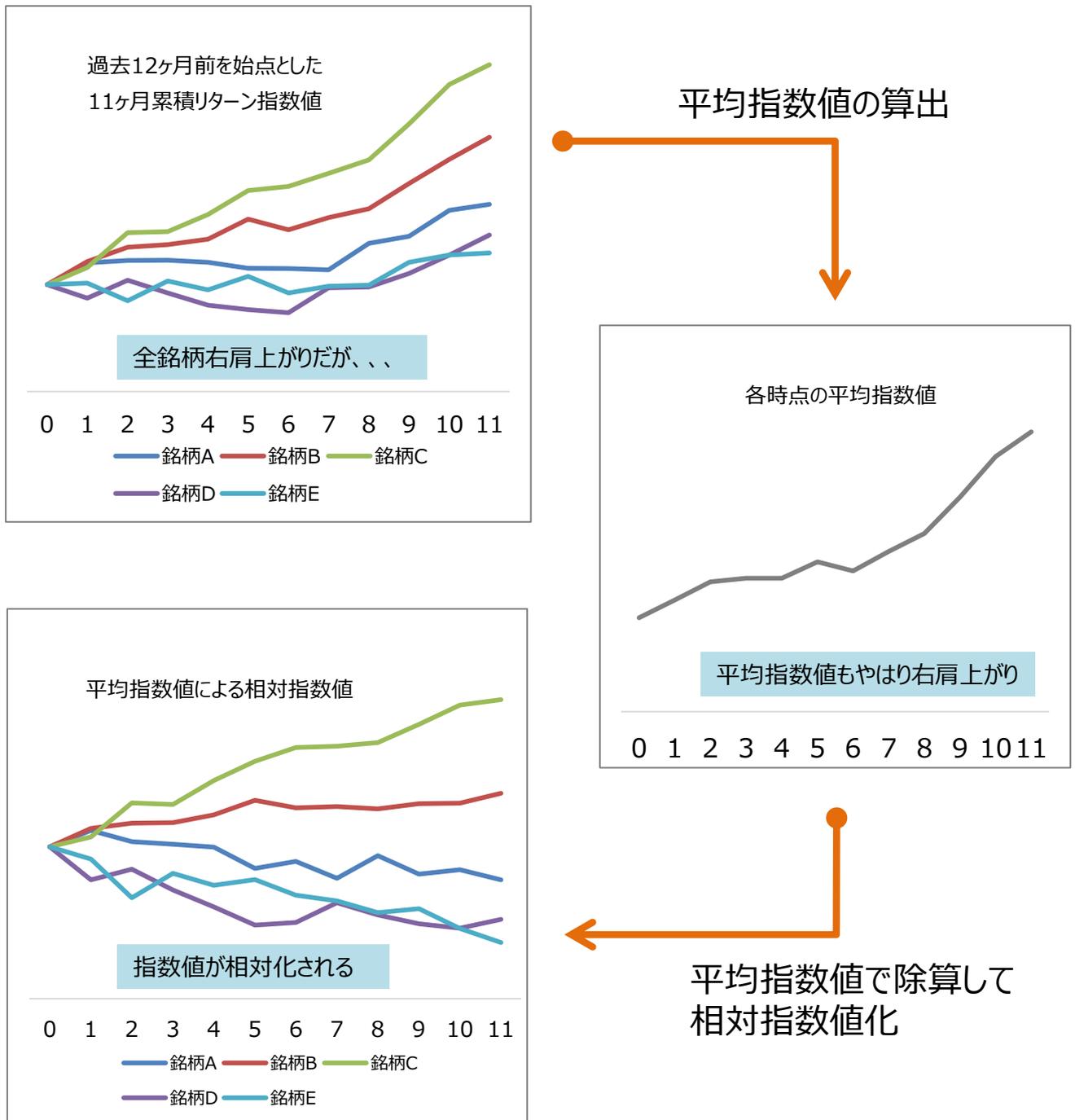


## マーケットの影響（ベータ）を除外して相対化する

### 手順4：相対指数値化

今回は、各銘柄の指数値を相対指数値化します。図6の左上グラフを見ると、全銘柄の指数値が上昇しているように見えるかと思いますが、これはおそらくマーケットのベータ（日本株ならTOPIX等の指数値そのものの動き）の影響が大きく出ていると思われます。このベータの動きを除外した上で、マーケットの中で相対的に上昇、下落を判別することを相対化、とここでは定義します。相対化のために、11か月指数値の各時点における平均指数値を作成し、全ての銘柄の指数値をこの平均指数値で割り算します。この作業により相対指数値を作成し、同じ時点における銘柄間の比較を可能にします。

図6.相対指数値化までの流れ



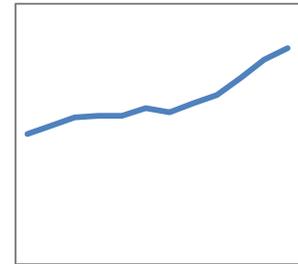
●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

## グラフ化とクロスセクション分析

### 手順5：グラフ化

今回は画像認識処理を取り扱うので相対指数値を画像化します。手順4で作成した相対指数値情報を、png形式の画像に変換します。この際、指数値は折れ線グラフを採用しました。また、縦軸の上下限の取り方ですが、今回は手順4で作成した相対指数値の、クロスセクション方向※での最大値と最小値を算出し、基準である1からの乖離が大きい方を採用しました。この値を上下限に適用することで、中心が基準かつ始点の値である1.0となるようにしています。また、通常のグラフには横軸、縦軸やラベル、グラフタイトル等がありますが、それらは全て非表示とし、純粋に折れ線グラフのみが表示される状態にしています。なお、ピクセル数は縦20×横20としています。折れ線グラフは単色ですが、情報としてはRGBを持たせています。

図7.最終的なグラフ  
(実際には四角枠も無い)



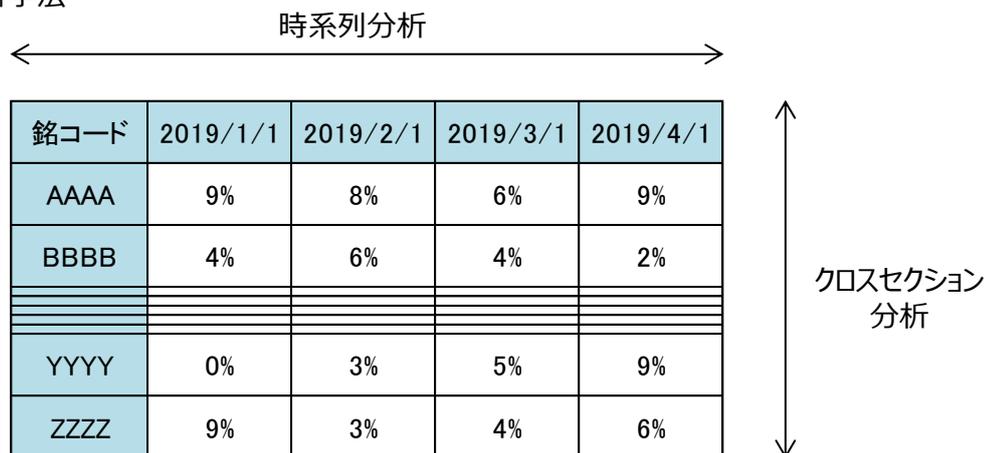
### 手順6：クロスセクション（月次）で画像データを作成

上記の手順1～手順6を毎月繰り返すことで、画像データを収集しています。最終的には、約35万枚の画像データを準備しました。

#### ※（参考）クロスセクション分析とは？

今回は一般的にクオンツ分析で活用されるクロスセクション分析をします。ある一時点における、銘柄間の関係性（相対感等）を分析する作業です。図8をご覧ください。縦軸方向にデータを眺めるのが、クロスセクション分析です。例えば、2019年3月時点における、全銘柄のモメンタムファクター値の関係性を分析する、ということになります。また、ある銘柄における、2019年3月と2019年2月の値の比較をする、という作業は時系列分析となります。今回の分析では、各時点における累積リターン値を相対化したチャート画像になっています。後で出てきますが、この画像データを過去10年間分学習させますが、この過程では時点間での比較作業は入っていませんので、時系列の変化を扱っている訳ではありません。

図8.分析手法



### 手順7：出力データの準備

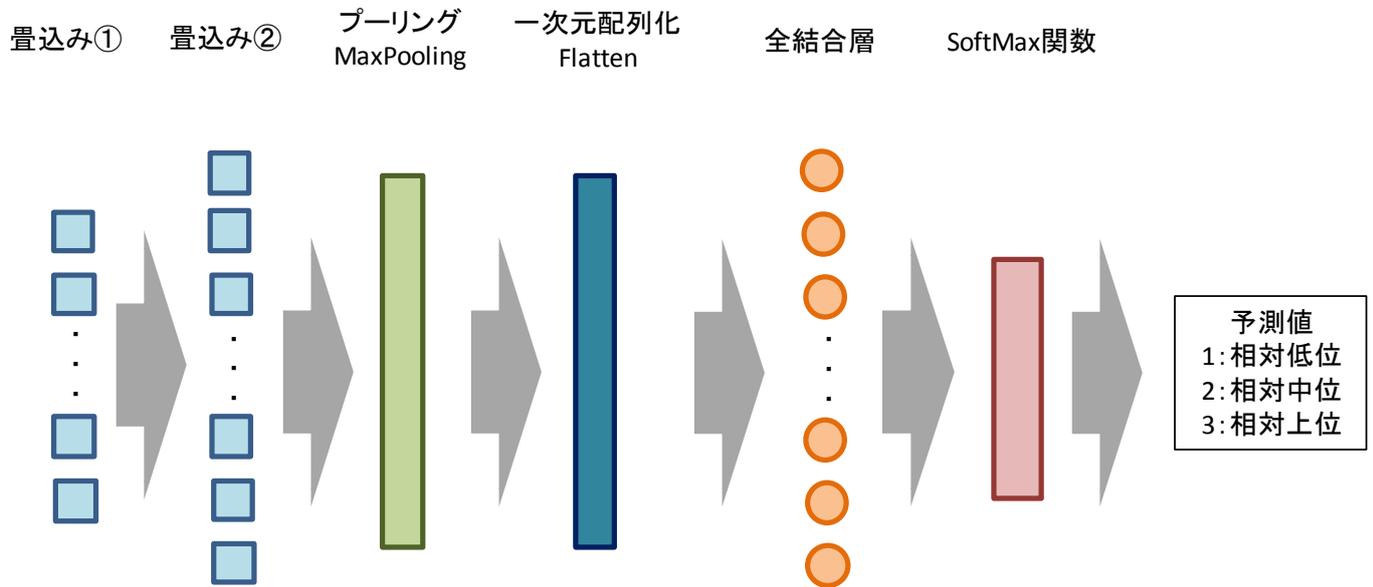
こちらは機械学習モデルによる学習させたい情報、言い換えると予測させたい情報になります。手順1で取得した13か月目の月次リターンが対象になりますが、今回は分類モデルを作成します。そこで、毎月、月次リターンについて、上位1/3はラベル3、中位1/3はラベル2、下位1/3はラベル1となるように加工しています。こちらのデータの外れ値処理は実施していません。

## 画像認識モデルの概要

### 3. 機械学習モデルの作成

今回は画像認識のためのモデルとして、CNN（畳み込みニューラルネットワーク）モデルを利用します。モデルの概要は以下であり、今回はあまり複雑にならないシンプルな構成を採用しています。

図9.CNNモデルの概要



### 4. シミュレーション結果

2000年2月から2010年1月の10年間を学習期間とします。この期間における画像枚数は約17万枚です。この学習済みモデルで評価するのは、2010年2月から2019年3月の10年間になります。この期間の画像枚数は約18万枚です。

なお、乱数による結果の再現性については、乱数シードは固定せず、代わりに10回試行しています。後述の正答率及びリターン値は、毎月における各試行の中央値を採用しています。以下が予測されるラベルの正答率推移になります。3分位ラベルの数はほぼ均等なので、ベースの正答率は33.3%となります。時系列で推移を見ると、概ね33.3%は超えているかと思えます。

図10.全体の正答率の時系列推移



●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

## シミュレーション結果

では、ラベル別の正答率を見てみましょう。図11を見てみると、傾向として中位ラベルは当てられています。下位と上位予想、特に上位予想ラベルの正答率は低位にとどまっています。

ラベルの正答率単体は芳しくありませんでしたが、一方で投資としての効果を測定する上で重要である、リターン値はどうでしょうか？ 予測期間における、各ラベル予想銘柄のリターンを累積した結果が図12になります。各ラベルにおける銘柄のウェイトは等ウェイトとしています。まず、結果から分かることは、上位1/3予想銘柄の累積リターンがプラスであることから、どうやら画像認識はモメンタム効果の強化（エンハンス）策として一定の効果があるかもしれません。正答率は低かったことから、あまり上昇銘柄を当てられているわけではないですが、大きく上昇する銘柄は当てられている可能性があります。ただし、2014年の上半旬をピークに、そこからは緩やかにマイナスの収益になっています。これは、計測期間において、2014年上半旬までは今回の学習期間のデータで表現可能だったが、2014年上半旬以降は何らかの構造の変化があった可能性が考えられます（あくまで推測です）。

図11.各ラベルの予想正答率

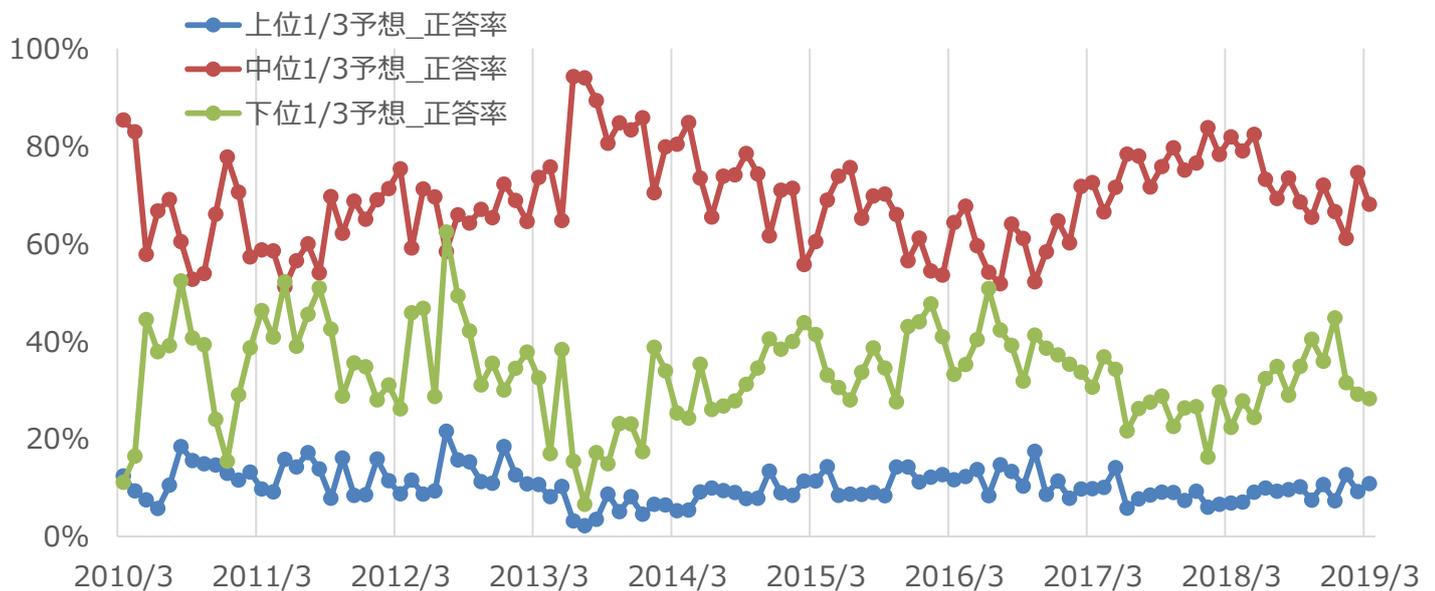
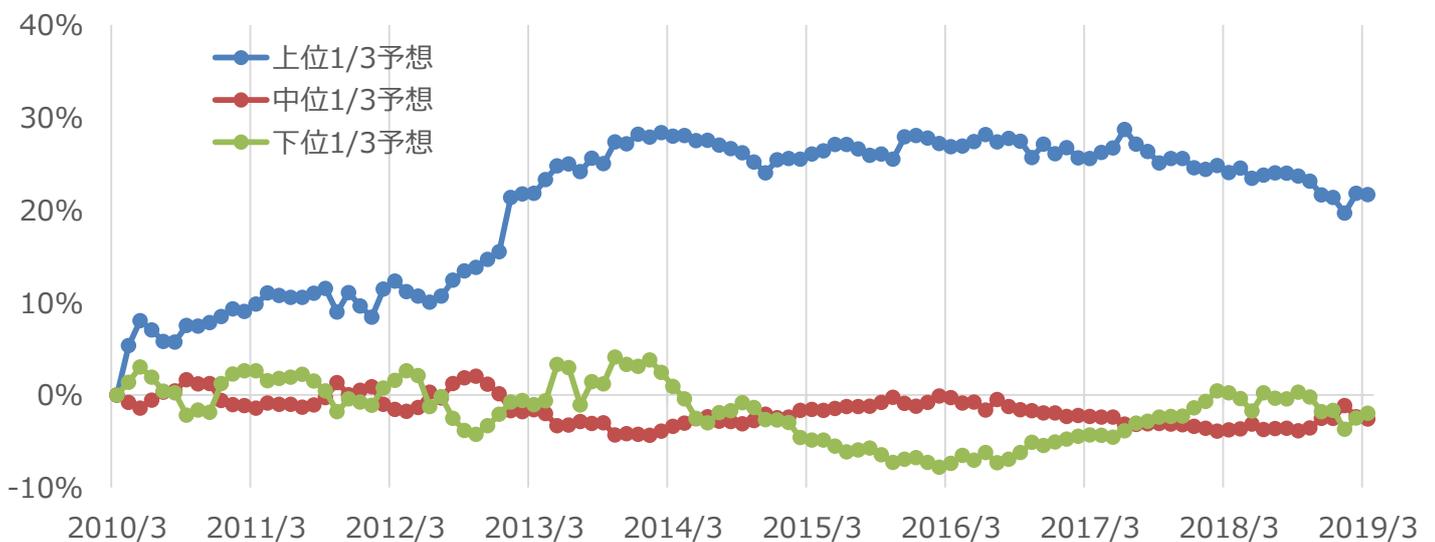


図12.各予想ラベル銘柄の平均累積リターン



●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。

## 今後のレポートテーマ

### 5. 改良方針

先ほどの結果を改良するにはどうしたら良いでしょうか？いろいろな手法が考えられますが、データに手を加えないのであれば、学習期間を毎月ローリングしていく方法が考えられます。例えば、学習期間を5年と決めて、毎月モデルを直近5年で再学習し、予測値を作成していくことが考えられます（Walk Forward方式）。また、データに手を加えるのであれば、月次の折れ線をより粒度の高い情報、例えば日次に変えること等が考えられます。

または、昨今の画像認識技術で良く活用される、転移学習を活用するのも手です。これは、ほかのデータセットで学習させたモデルを応用する手法です。大規模データで事前に学習されたモデルが既に様々な所から提供されています。このモデルをそのまま使うか、一部再学習（ファインチューニング）するのも有効な可能性があります。

### 6. 今後の展開

次回以降も、引き続きモメンタム効果に焦点を当てた分析を継続します。①今回の画像認識モデルの改良を試みる、②サイズやバリュー等とモメンタム効果の関連性を確認して、機械学習モデルに利用する新たな特徴量（入力データ）の作成を目指す、などを予定しています。

（筆者の都合でテーマが変わる場合があります。）

#### ～執筆者の紹介～

吉野貴晶（写真：右）

「日経ヴェリタス」アナリストランキングのクオンツ部門で16年連続で1位を獲得。ビッグデータやAIを使った運用モデルの開発から、身の回りの意外なデータを使った経済や株価予測まで、幅広く計量手法を駆使した分析や予測を行う。



高野幸太（写真：左）

ニッセイアセット入社後、ファンドのリスク管理、マクロリサーチ及びアセットアロケーション業務に従事。17年4月に投資工学開発室に異動後は、主に計量的手法やAIを応用した新たな投資戦略の開発を担当する。

●当資料は、市場環境に関する情報の提供を目的として、ニッセイアセットマネジメントが作成したものであり、特定の有価証券等の勧誘を目的とするものではありません。●当資料は、信頼できると考えられる情報に基づいて作成しておりますが、情報の正確性、完全性を保証するものではありません。●当資料のグラフ・数値等はあくまでも過去の実績であり、将来の投資収益を示唆あるいは保証するものではありません。また税金・手数料等を考慮しておりませんので、実質的な投資成果を示すものではありません。●当資料のいかなる内容も将来の市場環境の変動等を保証するものではありません。