〈財務省財務総合政策研究所「フィナンシャル・レビュー」令和元年第４号（通巻第 139 号）2019 年９月〉

# AI/フィンテックが変える資産運用

加藤　康之[*1]

要　約

資産運用でも IT 化が急速に進んでおり，AI やフィンテックが広く使われるようになっている。本稿では AI やフィンテックが資産運用手法の開発に応用されている実例を幅広く紹介している。応用されている技術の中核はテキスト情報を数値データ化するテキストマイニングであり，それはディープラーニングによって進化している。ビッグデータは運用モデルへの入力データ母集団を劇的に拡大し，これらデータをディープラーニングを使って解析して高度な予測モデルが開発されている。一方で，AI はモデルのブラックボックス化という弊害をもたらしておりその対応が課題となっている。また，AI による最適化技術やネット上のビッグデータからリスクファクターの推定という投資理論への貢献も試行されている。一方，フィンテックはウェルスマネジメントを自動化し，それがロボアドバイザーとして小口で投資未経験者層の資産運用普及に貢献している。また，最近注目度の高い ESG 投資でもビッグデータの応用が進んでいる。

キーワード：フィンテック，資産運用，ビッグデータ，ディープラーニング，テキストマイニング，LSTM，ロボアドバイザー，ウェルスマネジメント，ESG
JEL Classification：C1, C4, C6, C8, G1

## Ⅰ．AI/フィンテックの現状

新聞紙上の経済覧で AI やフィンテックの記事を見ない日はないと言っても過言ではない。まさに AI/フィンテックブームと言える。しかし，今回のブームは一過性ではなく，資産運用ビジネスにも大きな変革をもたらそうとしている。本節では本稿全体の導入部として今回の AI/フィンテックブームの特性と現状を概観する。

### Ⅰ－１．フィンテックの台頭と第３次 AI ブーム

日本銀行のホームページによれば，フィンテック（FinTech）とは，金融（Finance）と技術（Technology）を組み合わせた造語で，金融サービスと情報技術を結びつけたさまざまな革新的な動きを指す，とある。AI もフィンテックの一部と考えることが出来るが，AI だけに限れば，そのブームは今回が初めてではない。野村総合研究所（2016）によれば，過去に

＊1　株式会社お金のデザイン　お金のデザイン研究所所長，首都大学東京特任教授，京都大学客員教授

2回のブームがあり，今回は第3次ブームということになる。事務処理など通常のコンピュータ利用を超える人間の知能に迫る能力としてのAIが期待されたのは1980年代の第2次ブームが最初と言え，専門的な知識ベースとエキスパートシステムという専門分野に特化したAIが模索された。医療診断に代表される専門的な知識ベースを蓄積することに力を割いたが，応用範囲はきわめて狭く一般の実用には耐えなかった。過去のAIブームの特徴の1つは金融分野への応用があまり注目されなかったことである。当時のAI技術が金融に馴染まなかったこともあろうが，そもそも金融のIT化が十分に進んでいなかったことが最大の理由であろう。

今回のブームはAI/フィンテックブームであり，フィンテック企業の台頭によりIT，なかでもAIの金融実務への応用が進んでおり，フィンテックとAIが同時に注目されている。特に今回はインターネットの普及という観点で過去のブームと大きく異なっている。第2次AIブームの問題は何と言ってもデジタル化されたデータの圧倒的な不足であったと言える。当時，研究者や開発者は膨大な紙媒体から必要なデータを探し出しコンピュータに入力していたが，その入力だけで疲弊してしまっていた。第2次AIブーム以後に起こった極めて重要な技術的変化はインターネットである。インターネットの普及により多くの情報がデジタル化された。これにより使えるデータ量が劇的に増えた。AI開発に最も必要とされるのはデータでありインターネットの影響は計り知れない。特に実証分析を中心とする金融ではなおさらである。さらにインターネットはフィンテック企業を産み出した。ITスキルが高い新興のフィンテック企業群が積極的にAIを利用している。フィンテックというユーザーが付いたことで，ユーザーからAI開発サイドへのフィードバックが活発に起こり，AI開発を実用性の高い方向に導いている。また，既存の金融機関も人手不足の中，合理化の一環としてAIの導入を進めざるをえなくなっている。以上のような要因に加え，テキストマイニングやディープラーニングに代表されるAIの要素技術の進化があった。さらに，計算速度や記憶容量などのハードウェアの進歩も顕著である。

### Ⅰ－2．本稿で扱うAI/フィンテックの範囲

ところで，AIと言っても何をもってAIとするのかは人によって異なる。コンピュータの導入期にはコンピュータ自体がAIと考えられたであろう。今，パソコンのエクセルを使っていることでAIを使っていると考える人はいない。同じコンピュータの応用技術であってもAIであるためのハードルは技術の進歩とともに高くなって行く。参考までに，一般社団法人人工知能学会の設立趣意書の冒頭には次のような文章がある。「頭脳の働きに代わる機械が欲しいという人類の夢は，大量の数値データに対して複雑な計算を高速に行うという面では，電子計算機により実現された。現在の情報処理技術はこの意味においては，人間の能力をはるかに越えたものといえるが，一方，思考という本質的な面では，全くといっていいほど無力である。人工知能は大量の知識データに対して，高度な推論を的確に行うことを目指したものである。」つまり，AIでは単純な情報処理ではなく高度な推論を行うことが必要条件になる，ということである。第3次ブームでは，ディープラーニングに代表される高度な機械学習システムの利用に重点が置かれている。また，データに関しては，テキスト情報，画像情報，IOTで補足されるデータなどいわゆる非伝統的なデータ（オルタナティブデータ）を広く対象としている。なお，オルタナティブデータの導入によりデータ量は膨大になりビッグデータと呼ばれるようになった。膨大な量のデータを必要とするAIの利用においてビッグデータは今後ますます重要になろう。特に実証データへの依存度が高い金融では重要であり，今後さらに新しいデータの開発が進むだろう。

本稿では，高度な機械学習であるディープラーニングの利用を中心に想定しつつも，AI

やフィンテックの定義を明確に設けず，資産運用において先進的 IT の取組をしているケースを念頭に置いて考察する。なお，アルゴリズム取引などの短期トレーディングは含めない。

# Ⅱ．ディープラーニングと資産運用手法の開発

ここでは，AI が資産運用手法の開発業務でどのように使われるのかを考察する。特に，AI の進展とともに資産運用で利用が急増している AI の要素技術に着目する。つまり，テキストマイニングとディープラーニングという２つの要素技術である。具体的な応用先としては，テキストマイニングによるテキスト情報の数値データ化，動的資産配分モデル，運用スタイル評価を取り上げた後，AI 導入によるブラックボックス化という課題を指摘し，最後に運用機関へのインタビューから彼らの AI 対応状況を示す。

## Ⅱ－１．テキスト情報とテキストマイニングの活用

従来の資産運用手法では，GDP や金利などに代表されるマクロ経済データ，あるいは，個別銘柄の株価や企業財務データなどに代表されるミクロ経済データを用いて投資判断を行ったり，資産運用モデル開発を行ったりするのが一般的であった。これら数値化されたデータは項目名やデータの定義が明確にされ時系列データも整備されており「構造化」されたデータとも呼ばれる。構造化データは入手が容易であり，かつ，他市場（あるいは他銘柄）との比較や過去の値との比較が簡単に出来る。また，モデルへの組み込みなど取り扱いが容易であり資産運用では広く使われてきた。一方，これらのデータの課題としてはデータの種類が限られていること，データ作成の過程で情報が集約化されることが多いため重要な情報が欠落している可能性があること，そして，特にマクロ経済データや企業の財務データ等はその事象の発生からデータの公表までに１，２か月という長い時間がかかること等である。一方，ニュース，公的機関による公表資料，政府や企業幹部の記者会見，あるいは SNS 上のメッセージなどのテキスト情報はリアルタイムで大量に利用可能である。また，数値データに欠落している多くの情報を含んでいる。しかし，データ量が膨大なこと，取り扱いが難しいこと，あるいは，必要な情報の抽出が難しいこと等により，これらの膨大なテキスト情報も十分に利用されることは少なかったと言える。なお，これらのデータは構造化されていないので非構造化データとも呼ばれる。しかし，最近ではこれらのテキスト情報は多くが電子化され，デジタル的に読むことが可能になり一語一語利用することも出来るようになっている。さらに，テキストマイニング技術の進歩によりテキスト情報を正確にかつ情報の欠落を押さえながら数値化することも可能である。より広範でリアルタイム性の高いテキスト情報が数値化され既存の数値データと併せて資産運用に利用できるようになっている。

数値化の方法としては，テキスト情報から形態要素解析（テキスト情報を辞書に基づいて最小単位に分解して解析する方法）により単語を抽出し，その特性（例えば Good か Bad）に応じてスコアリングを行うというものが一般的である。このスコアに基づき数値化され市場動向の予測等が行われる。Tetlock（2007），Heston（2014），石島等（2013）は，単語の出現頻度やその Good/Bad を辞書で分類してニュース記事のセンチメントを評価し，株価との相関について論じている。また，取り出された単語が同時に使われる回数を示す共起頻度から単語を

グルーピングしてテキストを数値化することも行われている。グルーピングの方法としては，例えば，数値化された共起頻度データを主成分分析によりファクター化する。それらのファクターは時系列変数として使うことが出来る。和泉等（2010，2011）では，日銀が発表する金融経済月報を対象に形態要素解析しそれらを共起関係によりファクターに分類している。このファクターを既存の定量的モデルの変数として使えば，テキスト情報も組み入れたモデルが開発できる。京都大学の加藤研究室で行った陳＆孫（2017）の研究では，同様な方法で金融経済月報と日銀総裁記者会見のテキスト情報を数値変数化し既存の時系列モデルに組み込み，ボラティリティの予測に利用している。表１は金融経済月報で各年毎に抽出された単語の共起関係を頻度順に並べて上位のみを示したものである。第１列と第２列が共起した単語の組合せであり，第３列は２つの単語を接続したものである。第４列以降は各年に共起した回数を示している。ただし，単語は日経シソーラスの経済関連用語に当てはまるものに限定し，30 回以上観測された場合を抽出している。期間は 2003 年から 2013 年までである。

結果として金融経済月報では 607 組，総裁定例記者会見では 987 組が抽出された。全期間に対してこれらの共起単語を接続して一変数とみなして共起単語の出現パターンを結合した行列を作成する。この行列に対して主成分分析を行う。金融経済月報は上位 12 の主成分，総裁定例記者会見は上位 38 の主成分で累積寄与率は 60％に達する。それぞれの主成分スコアを計算し，それらを新たなファクターの時系列データと考えれば，これらのファクターを新たな外生変数として既存の時系列モデルに追加してモデルの精度を高めることが期待できる。

以上のように，デジタル化されたテキスト情報を解析し資産運用に応用することが可能になっており，今後大きな貢献が期待できる。ただし，これらの方法には課題も存在する。それは方法は単語ベースのアプローチであり，単語の意味はその使われる文脈によって大きく異

**表１　金融経済月報で抽出された単語の共起関係**

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 合計 |
| 2 | 金融 | 政策 | 金融政策 | 116 | 207 | 217 | 352 | 336 | 365 | 288 | 310 | 137 | 476 | 250 | 3054 |
| 3 | 金融 | 市場 | 金融市場 | 112 | 131 | 104 | 121 | 333 | 390 | 407 | 255 | 258 | 234 | 187 | 2532 |
| 4 | 金融 | 経済 | 金融経済 | 69 | 121 | 123 | 147 | 266 | 311 | 312 | 199 | 208 | 297 | 204 | 2257 |
| 5 | 物価 | 経済 | 物価経済 | 25 | 114 | 185 | 351 | 429 | 129 | 178 | 201 | 111 | 180 | 194 | 2097 |
| 6 | 金融 | 機関 | 金融機関 | 182 | 151 | 134 | 58 | 115 | 267 | 220 | 196 | 238 | 253 | 65 | 1879 |
| 7 | 日本 | 経済 | 日本経済 | 93 | 155 | 161 | 164 | 259 | 109 | 73 | 105 | 148 | 157 | 122 | 1546 |
| 8 | 経済 | 政策 | 経済政策 | 49 | 93 | 136 | 174 | 211 | 197 | 106 | 110 | 55 | 149 | 100 | 1380 |
| 9 | 中央 | 銀行 | 中央銀行 | 39 | 18 | 18 | 21 | 48 | 259 | 242 | 188 | 126 | 243 | 121 | 1323 |
| 10 | 世界 | 経済 | 世界経済 | 44 | 74 | 81 | 85 | 197 | 205 | 94 | 91 | 93 | 104 | 86 | 1154 |
| 11 | 日本銀行 | 金融 | 日本銀行金 | 42 | 43 | 61 | 40 | 22 | 73 | 143 | 140 | 143 | 282 | 136 | 1125 |
| 12 | 経済 | 市場 | 経済市場 | 26 | 77 | 85 | 103 | 240 | 127 | 80 | 54 | 94 | 71 | 75 | 1032 |
| 13 | 金融 | 物価 | 金融物価 | 10 | 51 | 58 | 107 | 102 | 90 | 79 | 98 | 68 | 181 | 175 | 1019 |
| 14 | 物価 | 政策 | 物価政策 | 24 | 60 | 93 | 140 | 137 | 109 | 57 | 85 | 39 | 159 | 100 | 1003 |
| 15 | 経済 | 情勢 | 経済情勢 | 35 | 63 | 85 | 167 | 173 | 64 | 95 | 59 | 64 | 50 | 58 | 913 |
| 16 | 米国 | 経済 | 米国経済 | 39 | 39 | 57 | 129 | 169 | 56 | 17 | 65 | 54 | 82 | 114 | 821 |
| 17 | 企業 | 金融 | 企業金融 | 49 | 65 | 17 | 26 | 48 | 48 | 285 | 64 | 60 | 118 | 26 | 806 |
| 18 | 金融 | 金利 | 金融金利 | 35 | 42 | 42 | 121 | 69 | 93 | 66 | 103 | 37 | 120 | 64 | 792 |
| 19 | 答 | 金融 | 答金融 | 74 | 53 | 52 | 52 | 91 | 78 | 107 | 65 | 48 | 73 | 63 | 756 |
| 20 | 金融 | リスク | 金融リスク | 33 | 48 | 20 | 41 | 85 | 127 | 111 | 94 | 67 | 68 | 51 | 745 |
| 21 | 経済 | リスク | 経済リスク | 12 | 26 | 38 | 94 | 115 | 100 | 62 | 73 | 71 | 64 | 89 | 744 |
| 22 | 物価 | 指数 | 物価指数 | 30 | 118 | 160 | 122 | 132 | 42 | 9 | 12 | 41 | 13 | 56 | 735 |
| 23 | 金融 | 状況 | 金融状況 | 41 | 57 | 40 | 45 | 87 | 110 | 100 | 45 | 55 | 78 | 49 | 707 |

2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2003-2013 | 試算1 | 試算2 | 2003-2013>=30

（出所）　陳紀洋，孫華君，2016,「テキスト情報を利用した金利期間構造のモデル化」，京都大学経営管理大学院

なってしまうという問題である。「円高」という単語は輸出企業にとっては Bad なニュースだが，輸入企業にとっては Good になる。つまり，単語だけでなくその文脈まで見て判断する必要がある。それを可能にしているのがディープラーニングなどの機械学習の利用である。

### Ⅱ－２．機械学習によるテキストマイニングの高度化

単語１つ１つではなく，テキストの文脈からそのセンチメントなどの情報を抽出するためには，単語だけではなくテキストとして理解する必要がある。Chen et al.（2015）によれば，文章から情報を抽出する方法には次の３つがあるとしている。すなわち，①既存の文章構造を利用した情報抽出，②教師なし機械学習，③教師あり機械学習，である。①は簡単で分かりやすい方法だがあらかじめ想定している構造に依存しており，限られたテキストにしか適用できない。②は単語の出現頻度や確信度の水準などテキストの統計的な特性から学習してその重要性を判断する。教師データは必要としない。③は既存のテキストを教師データとして準備しその教師データから学習して分類する。以上の中では，機械学習を利用し応用性の高い②と③が実用的であるが，開発の容易さでは教師データが必要ない②，精度では教師データを使う③が高いと考えられる。なお，単語ベースではあるが前節で紹介した陳＆孫（2017）は教師データを使っておらず②に該当すること言うことが出来る。

最近は多くのテキスト情報がデジタル化されているため，教師データの入手が容易になっており，機械学習システムとしてディープラーニングを用いてテキスト情報からセンチメントなどの情報を抽出する方法が注目されている。ディープラーニングを用いてテキスト情報からそのままセンチメントを抽出して株価との関係を分析している研究には，Chen（2013），Heston（2016），山本＆松尾（2016），五島＆高橋（2016）などがある。ディープラーニングを使うことによって文章の特徴量を抽出することが出来るが，ディープラーニングにも向き不向きがある。ディープラーニングがよく用いられる画像情報の認識には Convolution Neural Network（CNN；畳み込みニューラルネット）が最も良く使われていることは良く知られている。CNN ではニューラルネットに畳み込み層が導入され画像認識の機能を高めている。しかし，テキスト情報は画像のように２次元ではないため，必ずしも向いているとは言えない。代わりに Long Short Term Memory（LSTM） ユニットを持つ Recurrent Neural Network（RNN）が使われることがよくある。LSTM-RNN は時系列データのような連続した系列をニューラルネットワークで解析する手法であり，長さが異なる系列への適用も可能であるため，自然言語の解析に向いている（山本＆松尾（2016））。前の層の中間層を次の層の入力ベクトルの一部として用いるため，過去の系列を記憶することができる。文章データは可変長の時系列データであり，中間層の値を再び隠れ層に入力するというネットワーク構造は可変長のテキストの解析に便利である。山本＆松尾（2016）では，教師ありの LSTM-RNN を利用して，内閣府が公表している景気ウォッチャー調査から独自の景気指標を構築している。景気ウォッチャー調査では，家計動向，企業動向，雇用等代表的な経済活動項目の動向を敏感に反映する現象を観察できる業種の中から調査対象（景気ウォッチャー）を選定する。そして，彼らに景気判断（良いから悪いまでの５段階の数値情報）とその理由（テキスト）を聞き，まとめたものである。この景気判断と理由の組合せが教師データとして使える。これらの教師データをディープラーニングに学習させる。学習したシステムを使ってネット上で公開されているテキスト情報を解析し，景気判断（数値データ）に変換している。この研究では，内閣府の発表する月例経済報告および日銀の発表する金融経済月報などのテキスト情報を解析してセンチメント指数を作っている。この指数は日銀短観や景気ウォッチャー指数と高い相関があったことが報告されており，その有効性を示している。この結果を前提とすれ

ば，サンプリングが基本である既存のマクロ経済統計データに比べ，テキスト情報を利用して網羅性を高め補足率の高い変数が構築できるようになる。さらに，例えばある特定のセクターに限定してネット上のテキスト情報を解析して，そのセクターに特化したセンチメント指数を作ることが出来る。これらの指数によって，ある目的や業種に限定された数値指標を自由に，つまり公的機関の公表データの有無にかかわらず，自分で作ることが出来ることを示唆している。変数の多様化とモデル開発の自由度が大きく高まることが期待される。

以上のように，テキスト情報を高い精度で様々な目的に沿って数値変数に変換することが可能になっており実務への利用も始まっている。経済産業省は同様な技術を用いて経済指標（SNS × AI 景況感指数（ウォッチャー AI)等）を試験的に発表している。

### Ⅱ－３．機械学習による資産運用モデルの開発

一般的に資産運用モデルを開発するということは非説明変数と説明変数の間の関係を見つけ出すことだと言える。非説明変数は予測したい変数であり，説明変数は予測するために使う変数である。多くの場合，この関係は線型関係が仮定されており線型回帰分析がその主要な分析手法として使われてきた。しかし，回帰分析では変数が正規分布である等の厳しい条件がある。また，線型モデルはその構造が線型というシンプルなものであり，開発者がイメージ出来る範囲に限定される。ところが，実際のデータには非正規性があるものが多く，モデルの構造はより複雑でイメージの沸きにくい非線型な場合が多い。そこで，その非線型性をモデル化するためにニューラルネットなど高度な機械学習システムが利用されている。そして，さらに複雑な関係を見つけるために多層化したディープラーニングが用いられる。

前節ではテキストマイニングにより膨大なテキスト情報を数値変数化する方法について概観した。以下，これら多様な数値変数を機械学習システムを使って資産運用モデルに組み込む方法について検討する。ここでは，例として京都大学の加藤研究室の研究成果からボラティリティ予測モデルの開発について崔（2017）の研究を通してその方法を概観する。

ボラティリティの予測では具体的な数値そのものよりは，水準が高いか低いか，あるいは，直前より高くなるか低くなるかが重要である。そこで，ここではボラティリティの予測対象をボラティリティ水準の数値そのものではなくボラティリティ水準のレジームとする。そこで過去のリターン時系列から隠れマルコフモデル（直接観測されない状態をもつマルコフ過程）によってボラティリティ水準を４つのレジームに分類した上で予測を行う。対象は米国の株価指数である S&P500 と日本の株価指数である日経平均株価（以下日経平均）とする。機械学習アルゴリズムの比較も行うため，複数のアルゴリズムを用いて予測する。説明変数はトムソンロイター社が提供する市場センチメント指数である TRMI（Thomson Reuters MarketPsych Indices）を利用する。つまり，TRMI を説明変数としてボラティリティレジームを予測するというモデルになる。TRMI は，Thomson Reuter 社と MarketPhych Data 社が共同で開発したビッグデータ型のセンチメント指数である。ネット上の大手ニュースソースからソーシャルメディアまでをカバーして収集された日々2 百万件以上のテキスト情報をテキストマイニングによって解析して指数化している。TRMI には各資産別（S&P500，日経平均，石油価格，円ドルレートなど）に対応するインデックスがあり，また，インデックスの種類もそれぞれの資産に対して楽観指数（Optimism Index），悲観指数（Gloom Index），恐怖指数（Fear Index）など多岐にわたっている。テキストマイニングと AI 技術を駆使して開発された数値データと言える。S&P500 と日経平均のボラティリティ予測には表 2 にあるように当該資産に加え関連資産の TRMI も含めて複数の TRMI を説明変数として使った。

**表２　ボラティリティ予測に使う説明変数**

S&P500

| Asset | TRMI indices |
|---|---|
| S&P500 | sentiment |
| | gloom |
| | fear |
| Crude-Oil | sentiment |
| | gloom |
| | fear |
| Gold | sentiment |
| | gloom |
| USD Exchange Rate | sentiment |
| | gloom |
| EUR Exchange Rate | sentiment |
| | gloom |
| EmergingAsia (MSCI-50) | sentiment |
| | gloom |

Nikkei225

| Asset | TRMI indices |
|---|---|
| Nikkei225 | optimism |
| | fear |
| | stress |
| | trust |
| JPY Exchange Rate | gloom |
| | fear |
| Crude-Oil | sentiment |
| | fear |
| Gold | sentiment |
| | fear |
| S&P500 | sentiment |
| | gloom |
| Emerging Asia (MSCI-50) | sentiment |
| | gloom |
| | fear |

（出所）　崔盛旭，2017，「TRMI データを使ったレジームの分類」，京都大学経営管理大学院

これらの入力変数でボラティリティとの関係を学習させ予測に利用する。機械学習のアルゴリズムとしては，よく使われている次のものを使う。

・勾配ブースティング（Gradient Boosting）
・ニューラルネット
・畳み込みニューラルネット７層（CNN；Convolutional Neural Network）
・畳み込みニューラルネット 16 層（同）

勾配ブースティングは複数の弱学習器（真の分類と相関が低い分類器）を組み合わせるアンサンブル学習（複数のモデルを融合させて１つの学習モデルを生成する手法）の一つで逐次的に弱学習器を構築していく方法である。ステップ毎に弱学習器を構築して損失関数を最小化する。前のステップで誤って分類された例にはウェイトを高くして，次のステップで誤分類を避けうまく識別できるようにする。最も基本的で広く使われているシステムである。ニューラルネットは入力層，出力層に加え，一つの中間層（隠れ層）からなる構造を使う。畳み込みニューラルネットは画像認識によく使われるニューラルネットである。画像から特徴またはパターンを抽出していき活性化関数（複数の入力値を変換する関数）によって出力層の確率を計算する。隠れ層は複数の層を持っており，ディープラーニングになる。本研究の例では畳み込みニューラルネットが得意とする画像処理を行っているわけではない。しかし，入力データであるセンチメントインデックスの時系列データは画像のように認識することが出来る。したがって，より長い期間にわたる大きなサイズのデータセットの特徴を抽出することが出来ると考えられる。ここでは隠れ層の数の異なる２つの畳み込みニューラルネットを使う。

モデルの学習およびボラティリティの予測は日々の TRMI データを使い，21 日を１か月として直前１か月のデータを使って１か月先のボラティリティを予測する。学習期間は 1998.05.13～2007.12.31 で，アウトサンプルのシミュレーション期間は 2008.01.04～2017.10.31 である。４つのアルゴリズム毎の結果が表３に示されている。表３では学習期間の正解率（過去データの適合率），および，アウトサンプルの正解率（予測精度）の双方が各アルゴリズムについて示さ

表３　ボラティリティ予測のパフォーマンス（正解率）

S&P500

| 正解率 | アルゴリズム | | | |
|---|---|---|---|---|
| | 勾配ブースティング | ニューラルネット | 畳み込みニューラルネット（7層） | 畳み込みニューラルネット（16層） |
| 学習期間 | 97.53% | 93.83% | 96.92% | 98.60% |
| アウトサンプル | 68.71% | 65.50% | 64.74% | 58.98% |

日経平均

| 正解率 | アルゴリズム | | | |
|---|---|---|---|---|
| | 勾配ブースティング | ニューラルネット | 畳み込みニューラルネット（7層） | 畳み込みニューラルネット（16層） |
| 学習期間 | 90.34% | 86.55% | 93.16% | 94.30% |
| アウトサンプル | 65.72% | 57.84% | 66.22% | 59.08% |

学習期間：1998.05.13 – 2007.12.31
アウトサンプルのシミュレーション期間：2008.01.04 – 2017.10.31

（出所）　崔盛旭，2017，「TRMI データを使ったレジームの分類」，京都大学経営管理大学院

れている。

学習期間の正解率については，層の数が多いCNN が最も高かった。これは予想通りであろう。つまり，最も多層化した機械学習システムが複雑な構造に適合出来るということである。一方，アウトサンプルでは必ずしも高くない。つまり，多層化することにより過去の適合率は高まっているが，それが将来の予測精度を約束するものではないということである。オーバーフィッティング（過学習）になってしまっている可能性がある。過学習とはデータに適合し過ぎてしまうことである。ノイズも含んだ過去のデータに適合し過ぎることによりモデルが真の構造から乖離してしまうことを言う。真の構造から乖離すれば予測力が低下してしまう。学習の段階で適合度が高いことが必ずしも高い予測力を示すとは限らない。これは，データ数に対する層の数など機械学習システムをデザインする上で重要な視点である。機械学習システムを適用する場合，モデルの構造を想定した上で学習段階での適合度とアウトサンプルでの予測力の関係を慎重に見極める必要がある。

次にこのボラティリティ予測モデルを使ってTAA（Tactical Asset Allocation）運用戦略をシミュレーションしてみる。TAA 運用戦略は，対象となる複数の資産クラスのウェイトをアクティブに変更する戦略である。TAA 戦略によってバイアンドホールド戦略（当初配分した後そのままにしておく戦略）や等ウェイト戦略等の固定的戦略に比べて高いリターンを追求することを目的とする。ここでは，機械学習で予測した各資産クラスのボラティリティをもとにウェイトを決める。対象資産クラスとして，米国株式（ステートストリート S&P500　ETF），日本株式（野村日経平均 ETF），新興国株式（バンガード FTSE-Emerging Market ETF），米国債券（ブラックロック Core US Aggregate ETF）の主要 4 資産の ETF を利用する。ここでは次の運用戦略を検証する。つまり，ボラティリティが高くなると予測された株式資産への配分を減らし，ボラティリティが低下すると予測された株式資産への配分を増やすという戦略である。残りのウェイトは米国債券に配分する。表 4 にボラティリティ予測に基づく株式資産への配分比率の増減ルールを示す。ただし，新興国株式のボラティリティ予測には米国株式の予

表４　予測ボラティリティに基づく TAA 戦略

| 予想ボラティリティの変化 | 配分比率 | | |
|---|---|---|---|
| | 日本株式 | 米国株式 | 新興国株式 |
| 30%以上 | 15% | 10% | 5% |
| 10%～30% | 20% | 20% | 10% |
| 0%～10% | 23% | 23% | 23% |
| ▲10%～0% | 27% | 27% | 25% |
| ▲30%～▲10% | 30% | 30% | 27% |
| ▲30%以下 | 35% | 35% | 30% |

（残りは米国債券に配分）

（出所）　崔盛旭，2017，「TRMI データを使ったレジームの分類」，京都大学経営管理大学院

表５　予想ボラティリティに基づく TAA 戦略のシミュレーション結果

| | アルゴリズム | | | | 等ウェイト |
|---|---|---|---|---|---|
| | 勾配ブースティング | ニューラルネット | 畳み込みニューラルネット（7層） | 畳み込みニューラルネット（16層） | |
| リターン | 8.40 | 6.76 | 6.56 | 7.21 | 6.68 |
| リスク | 11.30 | 11.12 | 11.39 | 10.92 | 14.62 |
| リターン/リスク | 0.74 | 0.61 | 0.58 | 0.66 | 0.46 |

（シミュレーション機関：2008.01.04 － 2017.10.31）

（出所）　崔盛旭，2017，「TRMI データを使ったレジームの分類」，京都大学経営管理大学院

測値を使う。

機械学習のアルゴリズムとしては先ほど示した４つである。ベンチマークとして４つの資産に当ウェイトで配分したバイアンドホールド戦略を使う。その結果が表５に示してある。リターンおよびリターン／リスクで評価すると，この例では勾配ブースティングの成績が良かったが，以上はあくまでも単純なモデル開発例であり実用ではより精緻なモデル化が必要であることは言うまでもない。

以上，ボラティリティの予測を例に資産運用モデルに AI を利用した例を概観した。ボラティリティの予測以外にも資産運用モデルへの応用範囲は広く，今後さらに拡大すると予想される。伝統的なクオンツ運用を吸収しつつ，AI 運用という領域の範囲が拡大していくだろう。

なお，ここで取り上げた資産運用モデルはボラティリティの予測に軸足がおかれており，ポートフォリオは各資産へのあらかじめ決められたウェイトに配分するという簡便な方法で構築されていた。しかし，最終的に優れたポートフォリオを構築するためには，推定されたパラメータをもとに各資産クラスへのウェイトを最適化する必要がある。最適化のプロセスには，そのリスク構造を推定して何らかのアルゴリズムを導入する必要がある。ポートフォリオ最適化では，伝統的に平均分散法が用いられてきたが，資産運用の多様化や AI の進展により，様々な方法が提案されている。これについては，ウェルスマネジメントの節で触れる。

### Ⅱ－４．AI と運用スタイル評価

運用を外部機関に委託する年金基金にとって重要な資産運用プロセスの一つがファンドの運用スタイル評価ということになる。これまで，運用スタイル評価では伝統的な３ファクターモデル（Fama & French（1993））あるいはそれに基づくスタイルインデックスを利用した評価

が中心である。この方法では運用するポートフォリオを取り上げ，そのリターンの時系列特性やファンダメンタルズの断面特性をファクターの特性と比較することにより評価することが中心であった。しかし，近年，バリューファクターリターンの低下や低ボラティリティファクター等新しいファクターに対する注目が高まるなどファクターそのものの再評価についても議論されている。一方，新しい運用スタイル評価の枠組みについての研究も進んでいる。ここでは，スタイル評価の新しいアプローチとしてディープラーニングを使う方法を紹介する。

運用機関の日々の取引データを入力変数として，ディープラーニングを使って運用機関のスタイルを分析評価するシステムが試作され，報告書が発表されている。（ソニーコンピュータサイエンス研究所（2018），年金積立金管理運用独立行政法人からの受託研究）この報告書をもとに以下概観する。

年金基金など大手機関投資家は日本株式だけでも多くのアクティブ運用機関を採用している。これらの運用機関は多様な運用スタイルを持っておりスタイル分散されているが，同じスタイルと言われていても運用機関によってはそのパフォーマンス特性は異なる場合が少なくない。また，時間とともにスタイルが変化する可能性もあり常に注意深く監視しておく必要がある。一方，ディープラーニングやビッグデータに代表される最新の AI 技術を活用したファンドも登場し始めており，こうした新しいタイプのスタイル評価方法に関する課題もある。これらの理由により，スタイル評価方法の高度化が求められている。一般的にファンドは投資の基本哲学，戦略，手法等に基づいて類型化された資産運用の形態すなわち運用スタイルを持つ。そのスタイルによって各々のファンドが特徴づけられることになるが，端的にその違いが現れるのは日々の取引行動である。そこでソニーコンピュータサイエンス研究所では，取引データを元にファンドの運用スタイルを評価するシステム（運用スタイル分析器（Style Detector Array））を開発している。Style Detector Array はレファレンスとなる運用スタイルのそれぞれについての強度を評価する別個の識別器（以下，Detector）が並置された構成を取っており，それぞれの Detector はディープラーニングで実装されている。市場環境，企業の業績推移といった外部シナリオに加えて，ファンドの構成銘柄の内容（各時点での保有時価総額や評価損益）と日々の取引行動を時系列の入力として受け取り，教師データとしてあらかじめ用意した典型的な運用スタイルのそれぞれに対する類似強度を成分とした属性ベクトルとして，分析対象となるファンドの運用スタイルを出力する。なお，教師データとして，バリュー，モーメンタム，低ボラティリティなど典型的な 8 つの運用スタイルを模倣する仮想ファンドマネージャを設定し，シミュレーションにより生成した各スタイルの仮想ファンドマネージャの取引データを使って学習させる。これらのスタイルは既存のスタイル分類の域を出ていないため新しいスタイルの発見は難しいとい思われる。しかし，今後データの蓄積と研究が進めば新しい運用スタイル（つまり，ファクター）が発見される可能性もある。そうなれば，スタイル分析器としての付加価値が高まるだろう。このシステムはまだ実装されておらず実証結果を入手できないが，運用機関の評価・選択という基金にとって重要な分野に AI を応用した例として今後の進展が期待されている。

### Ⅱ－5．モデルのブラックボックス化と説明責任

ところで，資産運用モデルにディープラーニングを導入する際には留意すべき大きな課題がある。それは入力から出力に至るプロセスの背後にあるロジックやモデルの構造がどうなっているのかが直感的に分かりにくいという問題である。特に多層のディープラーニングでは，そのアルゴリズムの構造上の特性から入力（層）と出力（層）にのみ焦点が当たるが，その間のプロセスはブラックボックスになりがちである。これは，ディープラーニングの複雑な構造

上避けられないが，途中のプロセスが理解できないままでは，予測モデルとしての頑健性や経済的合理性に疑問をもたらすことになる。つまり，過去はうまく説明できたとしても，今後もそのモデルの精度が持続するのかが利用者には判断できないということである。特に市場構造が短期間で変化する可能性がある金融市場を対象にするモデルであるためそのモデルリスクが少なくない。そして，資産運用を受託業務として考えるとき，このブラックボックス化は受託者責任（フィデューシャリー・デューティ）を果たすのに困難をもたらす。つまり，AI モデルを使っている受託者が委託者に対して十分な説明が出来ないということである。ディープラーニングのように非線型に多層化されたアルゴリズムでは，なぜその結果が出てきたのかを直感的に説明することが簡単ではない。特にパフォーマンスが振るわないとき，説明責任を負う受託者にとっては，そのプロセスを説明することは不可欠である。ブラックボックス化でその出力の理由が開発者さえ分からない状況で委託者への説明は到底出来ないだろう。

この AI によるブラックボックス化の問題は資産運用のみならず，医療や法務など他の多くの分野でも重要な課題になっている。この課題を解決するために，入力と出力との関係を分かりやすく説明する方法の開発が進んでおり期待されている。この方法により AI は説明可能な AI（Explainable AI，略して XAI）となる。その方法論としては，モデルの構造を分かりやすくホワイトボックス化させる方法，そして，どんな入力があるとどんな出力があるのかというモデルの挙動を理解する方法の２つに大別出来る。ただし，モデルの構造は多様で複雑であり前者の方法は難しく，実際には後者の方法が実用化されている。Samek et al.（2017）は後者の方法として入力値に対する感応度分析（Sensitivity Analysis）と層別関連伝搬法（Layer-Wise Relevance Propagation）の２つを示している。感応度分析では，どの入力値の変化に対してどの出力値がどの程度変化するのかという変化量を調べることであり，入力値と出力値の間の関係を理解することである。なお，この方法ではモデルの構造には立ち入らない。一方，層別関連伝搬法では，出力層から入力層に逆にたどっていくことにより出力に貢献する入力を理解しようとするものであり，AI を層別に分割してシステム全体の挙動を理解しようとするものである。

機関投資家の資産運用において利用される AI は XAI である必要がある。それは，資産運用が説明責任を負っているからである。まだ新しい分野であるが，XAI への道は AI が資産運用で定着するためのきわめて重要な試金石になろう。

### Ⅱ－６．運用機関の AI 対応状況（インタビュー調査から）

日本でも主要な運用機関が AI を使った運用手法を開発している。現状について主要な運用機関へのインタビューを行った結果を報告する。以下，インタビューのまとめである。

① AI を使った商品，位置づけ

AI を使った運用商品としては個人投資家向け，機関投資家向けの双方が存在する。ただし，

- 商品名に「AI」を冠して AI を使っていることをアピールする場合，
- 要素技術としては AI を使っているが商品名に「AI」を使用せず既存の商品の位置づけのまま販売している場合，

の双方がある。この相違は当該運用商品のパフォーマンス全体における AI の貢献度の大きさ，および，各運用機関のマーケティング戦略に依存している。メディア等で AI に注目が当たっているため，特に個人投資家向けには AI の利用を明示してアピールする場合が多いようである。将来，AI の利用が通常の状況になれば AI をあえて明示することもなくなるだろう。

②利用されている AI 技術

利用されている AI 技術としては，テキスト

マイニングとディープラーニングが圧倒的に多い。証券会社の調査レポート，企業のディスクロ資料，行政の発表資料，ニュースなどの膨大なテキスト情報をテキストマイニングによって数値化し，新たな数値データを作り出すことが第一の応用である。テキストだけでなく，中央銀行総裁の表情や商業店舗の駐車場の混雑状況といった画像情報も利用されている。そして，それらの新たな数値データを含めて伝統的な数値データとともに膨大な量のデータをディープラーニングに入力してモデル化するという利用方法が中心である。なお，テキストマイニングの要素技術としてもディープラーニングが使われている場合がほとんどである。実証的な方法論に重きが置かれる資産運用では，AI による新たなデータ（オルタナティブデータ）の開発が中心になっていることを反映している。

③オルタナティブデータの活用

オルタナティブデータとしては，②で述べたようにテキスト情報が大半である。テキスト情報の範囲と量はきわめて多く数値データ化の可能性も多いため，当面この傾向が続くだろう。他に匿名化された商業店舗の POS データ，クレジット会社のデータなど，個人レベルの取引・信用データを集約化して作り出されるマクロデータも利用されるようになっている。なお，新しいデータ作りの優先順位は，超過リターンへの直接的な貢献度やモデルへの組み込みやすさなどが基準になっている。

④ AI 研究開発推進の組織，体制

多くの運用機関では，AI の研究開発や運用を担う組織は既存のクオンツ運用部門で行われている。これは，クオンツ系運用スタッフの専門性や学術的背景が AI 技術の習得に向いているという事情からと思われる。ただし，一部の機関では，研究開発を運用から切り離し別組織にして専業のスタッフを置いているところもある。別組織にする理由としては既存の考え方に縛られない新しい発想で研究してもらうためというものが多かったが，AI 専門家の処遇のためという理由もあった。これらの研究専業組織では 3〜10 名程度の体制となっている。また，大学の研究者との共同研究やスタートアップ AI ベンチャー企業との提携も多い。金融機関の伝統的な保守性を考えれば，特にスタートアップ企業との提携が予想以上に多かったが，必要に迫られた状況が垣間見られる。

⑤まとめ

現状では，テキストマイニングによるテキスト情報の数値化やディープラーニングの利用によるモデル開発といった既存のクオンツ運用の延長線上にあるものが多い。これは，資産運用業界では AI 利用の歴史も浅く，また，新規の技術開発も途についたばかりである状況にあるためと思われる。また，AI 開発が既存のクオンツ運用部門の担当者によって兼任されていることが多いためでもあろう。専業の研究開発組織の設置やスタートアップの AI ベンチャー企業との提携にも積極的なところが多く，今後，既存の枠組みにない運用手法の登場が期待できよう。

## Ⅲ．AI と投資理論

AI やビッグデータという道具の進展が投資理論そのものにも影響を与えていく可能性がある。それは，投資理論の多くが実証分析を土台にしており，ビッグデータは実証分析で使えるデータの範囲を拡大し，また，AI は実証分析の方法に革新をもたらすからである。以下では，

投資理論におけるリスクファクターの役割を解説した後で，AIやビッグデータを利用したリスクファクターの推定について具体的な2つの方法について取り上げる。

### Ⅲ－1．投資理論とリスクファクター

投資理論で最も重要なテーマの1つはリスクプレミアムをもたらすリスクファクターの推定である。ファクターが体系的に明示された最初の投資理論はSharpe（1964）が提唱したCAPM理論であろう。CAPM理論ではマーケットの超過リターンを唯一のファクターとしている。つまり，資産の超過リターンは資産の対マーケット感応度であるベータ値のみに依存するという理論である。これに対し，Fama & French（1993）は，マーケットファクターでは説明できないリスクプレミアムが存在していることを実証的に示し，バリュー，サイズの2つのファクターの有効性を示した。そして，マーケットファクターと併せて3ファクターモデルを提唱した。さらに，Fama & French（2015）では，3ファクターにクオリティと資産成長の2つのファクターを付加し，5ファクターモデルを提唱している。今や機関投資家による資産運用のコアポートフォリオになっているマーケットインデックスファンドやETFはCAPM理論を背景に成長してきた。また，2000年代に入り多様なファクターへのエクスポージャを有するスマートベータが大きく注目を浴びるようになったが，加藤（2015a）はスマートベータ隆盛の背景には3ファクターモデルの存在があると指摘している。さらに，ポートフォリオの評価に欠かせないアセットプライシングモデルもファクターにより構成されている。このように，ファクターは今や資産運用ビジネス全体の根幹を担うようになっている。特にスマートベータ導入以降は，ファクターの開発が運用商品の開発と直結し，ファクター開発競争が激化している。Harvey et al.（2016）は同論文発表時点で過去9年間に論文等で164個のファクターが発表されていると報告している。もちろん，有意なファクターの数は少数に限られていると考えるのが自然である。ちなみに，大手の証券市場インデックス提供機関である，FTSE社，MSCI社は両社とも基本になるファクターインデックスを6つ（Momentum, Quality, Size, Value, Volatility, Yield）と分類しており，これらは広く機関投資家に受け入れられている。ただし，両社のファクターの定義は微妙に異なっている。例えばバリューファクターについて見ると，FTSE社ではCash-flow／Price，Earnings／Price，Sales／Priceを合成して使っているのに対しMSCIではForward Price／Earnings，Enterprise value／Operating cash flows，Price／Book valueを使っている。どんなバリューファクターの定義が最適なバリューファクターなのであろうか。Fama & French（1993）はバリューやサイズというファクターを定義した後の最終章に「これらのファクターの背後にはよりファンダメンタルなファクターがある」と指摘している。つまり，彼らが提唱した3つのファクターの背後には真のファクター存在していることを示唆している。これまで開発された多くのファクターはほとんどの場合，専門家によって経験的に導かれたものであり，それらは最適なファクターとは言えない。

ここでは，AIを使ってより真に近いファクターを推定する2つの試みを紹介する。1つは，遺伝的アルゴリズムによるファクターの抽出であり，既存のデータから最適な組合せを見つけようとする方法である。もう1つは，テキストマイニングによって市場データを集約して算出された市場センチメント指標を新たなファクターとして使う方法である。

### Ⅲ－2．遺伝的アルゴリズムによるリスクファクターの抽出

優れたリスクファクターとはどんなものだろうか。最も分かりやすい条件としては，統計的有意性の高いリスクプレミアムを提供するものであると言える。つまり，そのファクターによって構築された高ファクターポートフォリオ（当

該ファクター値の高い銘柄からなるポートフォリオ）と低ファクターポートフォリオ（ファクター値の低い銘柄からなるポートフォリオ）とのリターン格差（スプレッド）が歴史的に十分に大きく統計的に有意であるということある。一般的には，ファクターを見つけるときは経験的に有効そうな変数を適宜作りそれを使ってスプレッドを計算し統計的な検定を行うというものである。株式の場合は，利益率やROEなど企業の財務情報やPBRやPERなどのバリュエーション指標を使う場合が多い。しかし，企業の財務情報は多様であり，また，それをROE（利益÷株主資本）のような指標にすると，その組合せはさらに多くなる。また，複数の指標を加重平均して一つの指標を作るといったことを考えると，その候補の数は膨大なものになる。専門家が経験的に良さそうなものを考えてファクターを作るというアプローチを採るのは無理からぬことである。そこで，ここでは最適化に用いられるAI技術の一つである遺伝的アルゴリズムを利用して，膨大な数の組合せからより最適なファクターを見つける方法を紹介する。なお，以下の方法は京都大学加藤研究室での研究成果である高（2017）に拠っている。

その方法はまず利益，売上高，負債，株主資本など基本的な財務項目の母集団を選び出し，任意の2変数を選び除することにより指標化する。それら指標の母集団から1つ，2つ，そして，3つを選び出しそれぞれに対し任意のウェイトで加重平均して新しいファクターの候補とする。このファクターで分類した高ファクターポートフォリオと低ファクターポートフォリオとのリターンスプレッドを計算し検定する。ウェイトを適宜変更しながら以上のプロセスを繰り返し，統計的有意性の高いファクターを選び出す。なお，ウェイトを変更し最適な組合せを見つけるアルゴリズムとして遺伝的アルゴリズムを使う。

遺伝的アルゴリズム（Genetic Algorithm）

遺伝的アルゴリズムとは，AI技術の一分野であり，遺伝子を交配，突然変異させながら徐々に最適解に近づけていく最適化手法である。生物の進化の過程を利用して作られている。遺伝的アルゴリズムの長所は，設計は容易ながらどのような形式の問題に対しても応用ができる点が挙げられる。ここではファクターを構成する指標の種類や個数を変化させて計算を行うため，最適化問題そのものがどのような特性を持つかが明確でない。遺伝的アルゴリズムであれば，問題によっては解が定まらないといった事態を避けることができる。また今回は線形結合でファクターを設計するが，異なる形で設計した場合にも，同様の手法で公平に比較することができる。一方短所として，大域的収束が保証されないという点が挙げられる。また交配や突然変異のパラメータとして一般的な値が存在せず，結果が設計方法に依存するという欠点も存在する。

以下，高（2017）で行った具体的な方法である。財務項目の母集団として，バランスシートと損益計算書から主要な財務項目13個（売上高，設備投資，減価償却，純利益，営業利益，配当金，流動負債，固定負債，支払利息割引料，総資産，流動資産，株主資本，売上原価）を使う。企業のファンダメンタルズを表す指標が真のファクターと考え，株価を使った時価総額等のデータは用いない。したがって，ファクターモデルで良く使われているバリュー（株主資本÷時価総額），サイズファクター（時価総額），あるいはモーメンタムファクター（直近のリターン）はここでは候補としない。Fama & French（1993）も指摘しているように，株価変動の背景にある真のファンダメンタルズを発見したいからである。13個の中の2つを任意に選びその比をとり指標とする。これらの指標を，ファクターを構成する要素とする。同じ2種であっても分母，分子が逆であるものは異なる指標として考えると，要素の数は ${}_{13}P_2 = 156$ 種類となる。そして，次式のようにファクターを定義する。

$$\text{Factor}=\sum_{i=1}^{n} w_i f_i$$

ここでnはファクターを構成する要素の数であり，ここでは，n＝1，2，3とする。また$f_i$はi番目に選び出した要素の実数値，$w_i$は$f_i$へのウェイトであり，任意の実数をとる。すなわち，1～3種類の要素を線形に組み合わせたものをファクターとして定義している。非線形なファクターの形式も考慮に値するが，ウェイトの値が任意であるため表現の幅が広いこと，実用上は複雑なモデルは好まれないことからここでは線形の場合のみを考える。

上で定義したファクターの値を母集団の各銘柄についてそれぞれ計算し，得られた値に従って降順に銘柄をランキングする。このとき，上位にランキングされたものほどリターンが高く，下位にランキングされたものほどリターンが低くなっていれば，このファクター値はリターンの高低に正の相関があり，リスクプレミアムを有していると考えることができる。逆に上位にランキングされたものほどリターンが低く，下位にランキングされたものほどリターンが高くなっていれば，ファクター値はリターンに負の相関があり，負のリスクプレミアムを有していることになる。この場合，重みの係数の符号を逆転させることで，正の方向に同じだけ相関が存在するファクターにすることになる。ここでは，財務項目データは2000年12月から2015年12月まで，株式リターンは2001年4月から2016年3月までのデータを使う。分析対象の銘柄母集団は東証上場の主要な銘柄の内，期間中すべてのデータが揃う140銘柄である。まず2000年12月の財務項目データを元にしたファクター値を計算してランキングする。上位20％の銘柄を高ファクターポートフォリオ，下位20％の銘柄を低ファクターポートフォリオとして，それぞれ等ウェイトポートフォリオとして2001年4月から2002年3月までの月次リターンを計算する。そして，高ファクターポートフォリオと低ファクターポートフォリオのリターンスプレッドを計算する。以上の作業を2001年以降2016年まで毎年繰り返し行い，15年分のスプレッドリターンを計算する。このリターンスプレッドが15年間累計で最大になるような指標の組合せと各指標のウェイトを遺伝的アルゴリズムで探索する。

要素となる指標を２つ組み合わせた場合についての分析結果を表６に示す。結果を見ると，支払利息割引料や流動負債など過去の研究であまり使われてこなかった変数が上位に現れてい

**表６　遺伝的アルゴリズムによるファクターの探索（指標が２個の場合）**

| 係数1 | 指標1 | 係数2 | 指標2 | ﾘﾀｰﾝ差(%) |
|---|---|---|---|---|
| 0.67 | 売上高/営業利益 | -1.33 | 流動負債/支払利息割引料 | 0.796 |
| 0.462 | 支払利息割引料/総資産 | -1.538 | 営業利益/売上原価 | 0.782 |
| 0.298 | 売上高/営業利益 | -1.702 | 総資産/支払利息割引料 | 0.779 |
| -0.148 | 営業利益/売上高 | -1.852 | 総資産/支払利息割引料 | 0.778 |
| 0.064 | 売上高/総資産 | -1.936 | 営業利益/支払利息割引料 | 0.777 |
| -1.303 | 売上高/支払利息割引料 | 0.697 | 売上高/営業利益 | 0.767 |
| -0.339 | 営業利益/売上高 | -1.661 | 営業利益/支払利息割引料 | 0.764 |
| -0.557 | 設備投資/減価償却 | -1.443 | 総資産/支払利息割引料 | 0.759 |
| 0.474 | 減価償却/株主資本 | -1.526 | 営業利益/流動資産 | 0.755 |
| -0.209 | 営業利益/売上原価 | -1.791 | 総資産/支払利息割引料 | 0.754 |

（出所）　高正義，2016,「機械学習を用いたファンダメンタルファクターの探索」，京都大学経営管理大学院

る。さらに，係数にマイナスのものが多くある。経験的にファクターを決める場合，マイナスのファクターを利用する場合はほとんどない。それは，直感的なイメージが沸かないためである。それが機械的に選択した場合，マイナスの係数も多くあり，人間の直感を超える選択が出来ることを示しており興味深い。

これらのファクターに関しては得られた組合せから逆にそのファンダメンタルな意味を考えることは意義があるだろう。ただし，これらのファクターも企業の財務情報データのみを使っており，そのデータの範囲に制約を設けていることに注意する必要がある。次節では財務情報に依存しないファクターについて検討する。なお，参考までに，指標母集団にバリューファクター（時価総額／株主資本）を加えると，バリューファクターが最も有意なファクターとして選ばれる。良く指摘されることだが，日本市場におけるバリューファクターの強さがうかがえる。

### Ⅲ－３．市場センチメントとファクター

前節で紹介したように，すでに使われているファクターとしてはROEやバリューファクター等株価や企業の財務情報を利用したものが代表的である。それは，企業の財務情報から作られるファンダメンタルズ指標がその企業のリスクを代表し，それが株価に織り込まれリスクプレミアムをもたらすと考えられるからである。しかし，どんな財務情報が将来のファンダメンタルズを的確に予測できるのかは明確であるとは言えず，また，時代や環境とともに変化すると考えられ，財務情報から長期的に安定したファクターを見つけるのは易しくない。ところで，実際に株式を売買して株価に影響を与えるのは投資家個人である。つまり，最終的には，投資家がリスクと考えるものが株価に織り込まれファクターになると考えるのが自然である。しかし，投資家の数は膨大であり，多くの投資家がどう考えているのかということを判断することはこれまで極めて難しかった。しかし，情報のデジタル化とテキストマイニング技術の向上により，テキストデータを中心としたネット上の様々なデータから膨大な情報を集約し，個々の投資家がどう考えているのかを集約することが可能になりつつある。とすれば，その集約された情報がより市場の実勢を表すファクターになると考えることが出来る。そこで，この集約された情報にリスクプレミアムが存在しているのかどうかを検証する。なお，以下では，京都大学加藤研究室の研究成果に拠っている。

ネットから収集された投資家の考えを集約した情報であるが，ここでは，本稿でボラティリティの予測に利用したセンチメントインデックスであるTRMIを利用する。TRMIには多くのインデックスがあるが，ここでは楽観（optimum）インデックスと悲観（gloom）インデックスを取り上げ，これらのインデックスがファクターになっているかどうかを検証する。

検証方法は，通常のファクターの検証と同様にして行う。つまり，ファクターに対する各銘柄のエクスポージャ（回帰分析の係数）を計算しこれを各銘柄のファクター値とする。このファクター値の大きさで銘柄を順位付けし，その順位によって母集団を高ファクターポートフォリオと低ファクターポートフォリオの２つに分ける。そして，高ファクターポートフォリオのリターンから低ファクターポートフォリオのリターンを引いてこのスプレッドをファクターリターンとする。以上の手順を，楽観インデックスと悲観インデックスについて適用し，それぞれのファクターリターンを算出する。対象は米国株式S&P500母集団の中から期間中通してデータが存在する380銘柄である。データ期間は2001/1/3～2017/10/11（日次）を使う。エクスポージャの再計算とポートフォリオのリバランスは半年に１回行う。エクスポージャは半年間の日次データで各銘柄のリターンを各センチメントインデックスのリターンで回帰分析して推定する。２つのファクターリターンを５ファクターモデル（マーケット，バリュー，サイズ，クオリティ，モーメンタム；

ファクターインデックスのリターンはMSCI社のファクターインデックスを利用）でリスク調整した後の超過リターン（切片）を検定した。その結果，楽観ファクターには正の超過リターンが信頼水準 10％で有意に観測され，悲観ファクターには負の超過リターンが信頼水準５％で有意に観測された。多くの人の気持ちを集約したセンチメントインデックスが市場に影響を与えていることが分かるが，これら人々の気持ちが市場のリスクを反映している可能性があると言える。もちろん，世論は株価と同様に，短期的に，その時のメディアの報道やムードに大きく影響されるため注意が必要だが，投資家の心情を読み込み定量化して利用する方向は今後増えるだろう。

# Ⅳ．ロボアドバイザーとウェルスマネジメント

資産運用分野では，これまで限られた富裕層向けのパーソナルな資産運用サービスであったウェルスマネジメントの大衆化が進もうとしている。AI の応用により，このウェルスマネジメントがロボアドバイザーという名でフィンテック化され，低コストで広く一般に普及しようとしているのである。以下，パーソナル化へのアプローチ方法とポートフォリオ最適化へのフィンテックの応用について概観する。

## Ⅳ－１．資産運用のパーソナル化と機能的アプローチ

ウェルスマネジメントは富裕層の投資家向けに高度にパーソナル化されたサービスであるが，フィンテックの進展によって一般の投資家にも同様にパーソナル化したサービスが提供されるようになっている。このサービスはロボアドバイザーと呼ばれているが，高度に IT 化され低コストで各投資家のニーズに合わせたサービスを実現しているのが特徴である。そのロボアドバイザーを支える技術が AI やビッグデータである。以下，ロボアドバイザーで実際に活用されているパーソナル化の技術について概観する。なお，以下の例では主に加藤（2015b）を参考にしている。

伝統的な資産運用では，ポートフォリオの構築において国内株式，国内債券，外国株式，外国債券という資産クラスの分類を使うのが一般的である。これらの資産クラスを組合せてポートフォリオを構築することにより，目標とするリスク・リターンを実現しようとするものである。保有資産額が少なく資産運用の目的も明確になっていない若い投資家にとっては，各人が許容できるリスク水準の範囲内で最大限のリスクを取り最大のリターンをあげるという単純な最適化で十分であろう。したがって，資産クラスの分類も伝統的な資産クラスという単純なもので十分である。この時に最も良く使われる最適化法が平均分散法である。しかし，成熟化・多様化した現代の投資家，特に高齢投資家の運用目標は必ずしもリスク・リターンだけで表すことは出来ない。それは彼らの資産運用の目的が多様だからである。歳を重ねると，資産規模，家族構成，趣味，そして，ライフスタイルは各人によって異なり多様になるのが普通である。資産運用の目的がこれらの多様な人生をサポートするためにあると考えれば，資産運用の目的も多様になる。若いときには意識しなかった寿命という時間の制約も現実的になり投資期間の設定も多様化する。したがって，資産運用に当たっては，リスク・リターンという視点に加えて，多様な目的，あるいは，目的を達成させるための機能という視点が重要である。この多様な機能を同時に満たす最適化手法，つまり，パー

ソナル化についてAIやビッグデータの利用が重要になってくると考えられる。

ところで，機能と言えば，Merton & Bodie（1995）が金融サービスを「機能的視点」から再定義している。金融ビジネスを銀行，保険，証券といった伝統的な業態ではなく，どんなサービスを提供するのかという機能的視点で6つに分類している。本稿では，資産運用に焦点を絞って機能的視点を導入する。実は，この考え方はすでに先進的な基金では導入されている。例えば，年金基金が導入しているLDI（Liability Driven Investment；負債指向投資）は，負債ヘッジポートフォリオとリターン追求ポートフォリオというそれぞれが明確な機能を有する2つの資産クラスに分けた上で全体ポートフォリオを構築する運用手法である。また，米国カリフォルニア州公的年金（CalPERS）では，全体ポートフォリオを成長資産，インカム資産，インフレヘッジ資産，リアル資産に分類し，それぞれの機能を明確にしている。さらに，日本初のロボアドバイザーであるお金のデザイン社のサービスTHEOでは全体ポートフォリオを株式（成長），金利（インカム），インフレヘッジの3つのポートフォリオを使って管理している。このように，投資家の目的を意識した運用手法を資産運用の機能的アプローチ，機能を明確にしたポートフォリオを機能ポートフォリオと呼ぶことにする。

以下では，投資家の多様な目的に対応するための機能的アプローチの例として，3つの機能ポートフォリオ（成長ポートフォリオ，インカムポートフォリオ，インフレヘッジポートフォリオ）を導入する。成長ポートフォリオは成長を，インカムポートフォリオは安定したキャッシュフローを，そして，インフレヘッジポートフォリオは物価上昇に対するヘッジを目的とするものである。以下，これらの機能ポートフォリオへの配分比率を多様な目的に沿って最適化する方法について検討する。なお，機能の分類については前節で取り上げたリスクファクターがベースになっており，現状ではリスクファクターと同様に経験的に決めている。しかし，今後ロボアドバイザーの利用者が増え個人投資家の資産運用の多様な目的が蓄積されてくると，これらが学習データになるため機能の分類方法も進化すると期待される。

### Ⅳ－2．機能ポートフォリオへの最適資産配分とAHP

機能的アプローチにおけるポートフォリオ構築では3つの機能ポートフォリオへの配分比率をどう決めるのかという最適化の方法が問題となる。伝統的なポートフォリオ理論では最適化に平均分散法が用いられる。しかし，リターンとリスクのみを勘案する平均分散法は，多様な目的の下では，必ずしも適した方法とは言えない。例えば，上で取り上げた機能的アプローチの1つであるLDIでは，負債ヘッジポートフォリオとリターン追求ポートフォリオの配分比率を決めるのに平均分散法を用いることはないであろう。それは，負債ヘッジポートフォリオの配分比率はその年金基金の負債価値の関数になるはずだからである。このように，投資家ニーズが多様化する中でのポートフォリオの最適化には多目的な最適化が必要になり，様々な方法が提案されている。代表的なものに遺伝的アルゴリズム，粒子群最適化，階層分析法などがある。その中で，今後，個人の特性を補足するビッグデータの整備とともに応用が期待されているのが階層分析法（AHP：Analytic Hierarchy Process）である。以下，階層分析法を用いて機能ポートフォリオの最適資産配分を計算する方法を解説する。

#### 階層分析法（AHP：Analytic Hierarchy Process）

AHPはSaaty（1980）によって提案された意思決定法であり，多様な要素が影響する意思決定や個人の好みといった感覚的な要素の影響を受ける意思決定に用いられている。多様な要素や個人の考え方が影響する退職後の資産運用においてポートフォリオを最適化，つまり，パーソナル化することに適していると考えられる。とく

に，パラメータの推定には経験値が必要であり，これまで資産運用で利用されることはあまりなかった。しかし，個人の属性情報と投資特性の関係をビッグデータで補足できるようになり，その精度が高まると期待でき，資産運用への応用も増えると思われる。AHP の資産運用への応用に関してはこれまでに Satty & Rogers & Pell（1980），Bolster & Janjigian & Trah（1995），宮崎（2003），Le（2008），加藤（2015b）などが導入的な研究を行っている。本稿では，以下，加藤（2015b）をもとに AHP を使って３つの機能ポートフォリオの配分を決める例を示し，ビッグデータによる高度化の方法を検討する。

AHP による階層構造

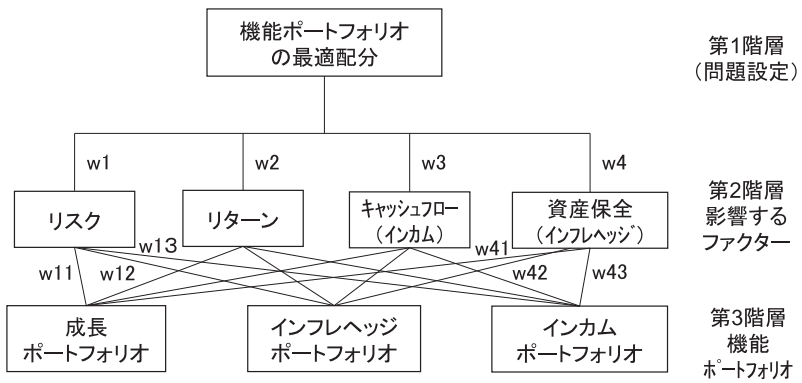
AHP では，問題解決のために必要な要素をグループ毎に階層構造にする。本例での全体構造を図１に示す。wij はある階層の要素 i における次の階層の要素 j に対する評価（ウェイト）である。ただし，第１階層の i=0 は省略されている。まず第１階層で与えられた問題（ここでは，３つの機能ポートフォリオの最適配分）を設定し，第２階層の４要素（ここでは，３つの機能ポートフォリオに関連した個人の志向を表すリスク，リターン，インカム，資産保全）と第３階層の３要素（３つの機能ポートフォリオ）によって構造化する。次に，各階層ですべての要素のペア（一対）を比較し相対値を与える。この相対値から各階層の要素に対する評価（ウェイト）を算出する。なお，階層の数はさらに増やすこともできる。

一対比較行列

各ペアの比較を行うのは，複数の評価軸があっても各ペアの相対値があれば１つの数値で比較出来るからである。一対比較の相対値は要素間のすべてのペアに対して与えられるため，一対比較行列として表現される。この例の場合，第２階層の要素は４つあるのでその一対比較行列は４×４の行列である。第３階層の要素は３つあるので一対比較行列は３×３の行列になるが，第３階層の行列は第２階層の各要素に対して別々に存在するので３×３の行列は４つある。ここでは，この比較に使う相対値を１から９までの値とその逆数とする。例えば，要素１が要素２に比べて絶対的に重要であれば，行列の（1，2）には９が割り当てられ，（2，1）には逆数の 1/9 が割り当てられる。なお，第２階層の行列は各投資家の特性や嗜好から投資家別に決め，第３階層の行列はすべての投資家に共通と考えて AHP の設計者が予め決めておく。

まず，第２階層の行列だが，ここでは A さんという仮想の投資家を考え，A さんの特性から第２階層の４×４の１対比較行列を作成する。A さんは 68 歳の退職者であり，保有資産

**図１　AHP による機能ポートフォリオの最適配分**



は3,000万円，労働収入はないものと仮定する。Aさんへの質問から年齢，資産規模，収入などの情報を取得し，これに既存の学習データを使って図2のような行列が推定されたとする。

この例では，リスクの行とリターンの列の交点に7という数字がある。つまり，Aさんにとって，リスクが低いということはリターンが高いということに比較して「かなり重要」だということが分かる。また，キャッシュフローと資産保全の交点は1になっており，これは2つの要素の重要度は同じであることを示している。なお，このAさんの1対比較行列の推定については，Aさんが質問に回答する際，質問を十分に理解できない場合やバイアスのある回答をしてしまう場合もあると思われる。したがって，サンプル数が増えると既存顧客の回答データとその後の資産運用データを教師データとして，新規顧客の質問回答時のバイアスを特定して修正することによってその人により適した行列の推定が可能になる。つまり，サンプル数が増えると一対比較行列というパラメータの推定精度が高まることが期待できる。

さらに，第2階層の4要素における第3階層の3つの要素（ここでは機能ポートフォリオ）間の一対比較行列を，過去リターンデータを教師データとし，また，専門家の経験的な判断を加えて推定し，図3のように推定したとする。

「リスク」という要素においては，インカムの行と成長の列の交点に9がある。これは，「リスク（が低い）」という要素に関しては，インカムポートフォリオは成長ポートフォリオに比べて「絶対的に重要」であることを示している。一方，「リターン」という要素では，成長の行とインカムの列の交点に7がある。これは，「リターン（が高い）」という要素に関しては，成長ポートフォリオはインカムポートフォリオに比べ「かなり重要」であることを示している。この第3階層の行列についても多くの顧客データが蓄積されれば，それらを教師データとして投資家の選好を反映した行列が推定できるようになる。以上により，階層構造のすべての一対比較ができたことになる。

固有ベクトルによるウェイトの計算

次に各階層で要素に対するウェイトwijを一対比較行列から計算する。これは行列の固有値から求めることが出来る。（詳細はSaaty（1980），加藤（2015b）等を参照）第2階層の一対比較行列（図2）で実際に固有値を計算しそれを基準化すると4つのウェイト（w1～w4）はリスク（48.5%），リターン（4.8%），キャッシュフロー（23.5%），資産保全（23.2%）

**図2　第2階層の一対比較行列**

Aさん（リスクの取れない退職者）の例

プロファイリングによってこの行列を推定する

| 後 / 前 | リスク | リターン | ｷｬｯｼｭﾌﾛｰ | 資産保全 |
|---|---|---|---|---|
| リスク | 1 | 7 | 3 | 2 |
| リターン | 1/7 | 1 | 1/7 | 1/5 |
| ｷｬｯｼｭﾌﾛｰ | 1/3 | 7 | 1 | 1 |
| 資産保全 | 1/2 | 5 | 1 | 1 |

**図３　第３階層（代替案間）の一対比較行列**

| リスク | | | |
|---|---|---|---|
| | 成長 | インフレヘッジ | インカム |
| 成長 | 1 | 1/5 | 1/9 |
| インフレヘッジ | 5 | 1 | 1/3 |
| インカム | 9 | 3 | 1 |

| リターン | | | |
|---|---|---|---|
| | 成長 | インフレヘッジ | インカム |
| 成長 | 1 | 5 | 7 |
| インフレヘッジ | 1/5 | 1 | 3 |
| インカム | 1/7 | 1/3 | 1 |

| キャッシュフロー | | | |
|---|---|---|---|
| | 成長 | インフレヘッジ | インカム |
| 成長 | 1 | 3 | 1/5 |
| インフレヘッジ | 1/3 | 1 | 1/7 |
| インカム | 5 | 7 | 1 |

| 資産保全 | | | |
|---|---|---|---|
| | 成長 | インフレヘッジ | インカム |
| 成長 | 1 | 1/5 | 1/3 |
| インフレヘッジ | 5 | 1 | 3 |
| インカム | 3 | 1/3 | 1 |

となる。退職者でリスク回避的な A さんにとってリスクの低さは最も重要でありウェイトが最も高い。一方，労働収入がないので次にキャッシュフローは重要となる。インフレがあれば資産価値が減ってしまうことから資産を保全する必要もある。一方，高いリスクを取ってまでリターン求める必要はない。

次に，第２階層の４つの各要素から第３階層の３つの要素に対するウェイト（wi1，wi2，wi3；i = 1，2，3，4，i は第２階層の要素を示す）を第３階層の一対比較行列（図３）から上と同様に固有ベクトルを使って計算する。最後に，３つの要素（機能ポートフォリオ）の最終ウェイトを計算する。そのためには，上位階層から各要素のウェイトを積算する。その結果，成長（13.4%），インフレヘッジ（30.4%），インカム（56.2%）と計算され，これが A さんにとっての最適ポートフォリオとなる。インカムを中心とし，インフレヘッジにも留意する。成長にはあまりウェイトを与えずリスクを抑えている。以上のように AHP によって，個人の特性や志向を反映したポートフォリオを多機能・多目的のフレームワークで構築することができる。

AHP で使われる一対比較行列については，利用者（サンプル）が増えれば増えるほど各個人の特性や投資への選好の傾向が分かり，それらを教師データとして利用できるようになる。ビッグデータの蓄積とともに，各個人により適したポートフォリオ運用手法の推定が可能になると期待されている。

## Ⅴ．ビッグデータと ESG 評価

機関投資家の資産運用では ESG 投資が世界的に大きなテーマになっている。ただし，まだ未成熟な分野であり新しい技術の応用も期待されている。本節では，AI/フィンテックがこの

ESG 投資でどのように利用されているのかを概観する。

### Ⅴ－１．ESG 投資とその課題

ESG とは，環境(Environmental)，社会(Social)，ガバナンス（Governance）の３つの頭文字 E，S，G をつなげたものである。そして，ESG 投資とは，これら３つのファクターに対する企業の取組み状況に基づいて投資対象企業を選別する投資手法のことである。例えば，$CO_2$ の排出量削減など環境（E）に配慮した経営を行っている企業，女性役員の登用（女性の社会進出支援）など社会（S）に配慮した企業経営を行っている企業，そして，社外取締役の採用などガバナンス（G）を重視した経営を行っている企業に投資をする，といった投資手法になる。ところで，企業の ESG に対する取組状況の評価は専門の ESG 評価機関によって行われ，ESG レーティングが各企業に付されている。しかし，この ESG レーティングについては多くの課題が指摘されている。評価機関がレーティング付与時に参考にする各企業の ESG 情報は企業の統合報告書などで開示されている。しかし，ESG 情報の開示はまだ始まったばかりであり統合報告書を発表している企業も限られている。情報量も十分であるという状況ではない。また，ESG 情報は非財務情報と言われ，定量化されていない情報が多い。したがって，ある程度定性的な判断が行われるため，評価機関が違うと同じ企業に対しても ESG レーティングが大きく異なる場合がある。GPIF（2017）によれば，2017 年時点での FTSE 社と MSCI 社による ESG 評価を共通する企業母集団でクロスセクションのプロットをすると無相関に近くなることが示されている。定性的評価が入る限りその差を埋めることは易しくないであろう。また，統合報告書など企業の開示はその頻度が限られており，日々刻々変わる企業の ESG 対応状況をリアルタイムに反映するのは困難である。さらに，数値でない定性的情報では企業の発表はどうしても自己に都合の良い方にバイアスがかかる可能性がある。これら ESG 評価への不安は ESG 投資を行う上で大きな課題となっている。

### Ⅴ－２．ビッグデータと ESG 評価

ESG 評価においてビッグデータや AI を使って行うというアプローチが始まっている。つまり，企業の開示情報に加えて，公的機関の情報，ネット上のニュース，消費者のコメントといった第三者の評価情報を大量に集めてビッグデータとして評価しようとするものである。石井（2018）によれば，ESG 評価におけるビッグデータや AI の具体的な応用としては，自然言語処理と分類アルゴリズムがある。自然言語処理は利用する情報の多くがテキスト情報であるため，その解析と意味の抽出ということになる。また，分類アルゴリズムは，テキスト情報等から抽出された意味を ESG レーティングを構成する要素に分類してスコア化することである。例えば，あるテキスト情報を解析し，それが企業の取締役会に関する情報でありガバナンス評価に関連する項目であるという分類を行う。そして，次にその分類された項目の評価を行いスコアを算出する。実際このような ESG 評価サービスを提供する企業も登場し始めている。

石井（2018）によれば，2013 年に米国で設立された TruValue Labs 社の場合，インターネット上の膨大なデータに着目している。インターネット上に存在する企業の ESG 関連ニュースを集約しスコアに変換している。あるニュースがその企業の ESG にとってポジティブかネガティブかというセンチメント評価を行う機械学習モデルがシステムの中核になっている。自然言語処理では機械学習モデルであるため教師データが機械に正しい答えを教え込む必要がある。TruValue Labs 社の場合，ESG の専門家が教師として正解をデータに与える。例えば，ニュースの記事がどのカテゴリーに対応するか，人間のアナリストが事前に評価を行い正解のデータセットを作成する。100～200 件のニュース記事に対し，まず教師が正解付けを行う。そして学

習したモデルに新しい記事を読み込ませ，カテゴリー分けをさせる。このステップで，機械学習モデルが正しくカテゴリーを識別するか検証する。正解との乖離（エラー値）を基に，モデルを更にチューニングし，モデルのパフォーマンスを上げていくとしている。処理するデータは，インターネット上で報道される各国の現地ニュース，国際的に報道されるニュース，業界誌，NGOやウォッチドッグ機関の報告書，そして様々なブログやソーシャルメディアを情報源としている。日次で約75,000件のニュースを読み込んでおり，月次ベースでは100万以上のデータ点をフィルターしている。ESGスコア自体は，2017年現在，約8,700社をカバーしており，ほぼリアルタイムでデータは更新されている。このビッグデータの大きさは高度な機械学習を可能にしている。また，伝統的な評価機関がカバーするデータや銘柄の数を圧倒している。AIやビッグデータの利用により同様なサービスを提供するESG評価機関が今後増加すると考えられている。

## Ⅵ．おわりに―AI/フィンテックと資産運用ビジネスの将来

### Ⅵ－１．AI/フィンテックを使えば儲かるのか？

本稿ではAI/フィンテックの資産運用への応用例を幅広く紹介した。AIの要素技術としてはテキストマイニングとディープラーニングが中心であった。テキストマイニングは膨大で未利用のテキスト情報を数値データ化し，新しいデータの開発に利用されている。また，膨大な種類のデータを使って複雑な構造をモデル化するのに力を発揮しているのがディープラーニングである。ところで，2018年8月30日付けの日本経済新聞によれば，国内で販売されているAI投信（AI技術を利用して運用されている投信）は10本あり，2017年3月以降のパフォーマンスを検証すると，そのほとんどが市場平均並みであるという。特に日本株式で運用されている投信に注目すると，市場平均に負けているものが多いとしている。AIを使えば超過リターンを得られるという簡単な話ではなさそうである。市場平均に対する超過リターンはゼロサムの競争である。そのため，市場参加者は他のプレーヤーの動向を常時モニターし戦略を変えてくることになる。せっかく高度なAIでモデル化された過去の構造もすぐに変わってしまうことになる。また，AIやフィンテックにより超過リターン獲得のコストが減少すれば超過リターン自体が減少すると考えることは自然である。この結果をもってAIは投資家にとってメリットがないと結論する論調もある。しかし，AIによる自動化で運用コストは下がり，また，市場の効率性が高まることが期待される。いずれにしろ，資産運用におけるAI利用の流れはもはや止まらない。AIを利用しないとそもそも競争に参加できなくなるということは容易に想像できる。現在，パソコンやインターネットを利用しないファンドマネージャはいないのと同じことである。また，本稿の後半で触れてきたように，AI/フィンテックの利用は，超過リターンを求める運用手法の開発だけではなく，運用の評価，新しい投資理論への貢献，パーソナル化された資産運用サービスの提供，新しいESG評価方法の開発などその裾野はきわめて広いことが分かる。そして，これらのAI/フィンテック技術を活用したフィンテック企業が台頭し，資産運用サービス利用者の裾野を広げ，個人の資産形成に貢献するようになるだろう。

### Ⅵ－２．資産運用における人間の役割は何か？

ところで，AI/フィンテック時代の資産運用

における人間とAIの役割分担はどうなるのだろうか。投資対象資産に関する情報を解析してそれをもとに投資判断を行い売買する，という「既存資産」の運用はますますAIに任せることになるだろう。この分野はモデル化が可能であり，人間がAI/フィンテックに勝てなくなる日はそう遠くないと思われる。では人間の役割は何だろうか。それは新たな投資機会の創造である。新しい価値の創造と言っても良いだろう。人間に求められる役割は，現在存在していない新たに富を産み出す投資資産を創出することである。アルファの創出からベータの創出への転換と言える。株式や債券もそれらが登場したときは新たに産み出された資産だったのである。不動産やインフラの証券化資産も同様である。当たり前のことだが，人間には創造力が求められるのである。

## 参 考 文 献


石井正太（2018）『ESG投資の理論―理論と実践の最前線―（加藤康之編著）第8章ESG評価の最前線：ビッグデータと人工知能の活用』一灯舎

石島博，數見拓朗，前田章（2013）「日次データを用いた市場センチメント・インデックスの構築と株価説明力の分析」『人工知能学会研究会資料』

和泉潔，後藤卓，松井藤五郎（2010）「テキスト情報による金融市場変動の要因分析」『人工知能学会論文誌』25巻3号，pp. 383-387

和泉潔，後藤卓，松井藤五郎（2011）「経済テキスト情報を用いた長期的な市場動向推定」『情報処理学会論文誌』52巻，12号，pp. 3309-3315

加藤康之（2015a）「株式投資の新潮流とスマートベーター―リスクプレミアムとベンチマークインデックス―」『証券アナリストジャーナル』2015. 5，pp. 53-64

加藤康之（2015b）『テクニカル詳細　高齢化時代の資産運用手法―キャッシュフロー管理と機能的アプローチ』一灯舎

高正義（2016）「機械学習を用いたファンダメンタルファクターの探索」京都大学経営管理大学院

五島圭一，高橋大志（2016）「ニュースと株価に関する実証分析―ディープラーニングによるニュース記事の評判分析―」『証券アナリストジャーナル』2016. 3，pp 76-86

ソニーコンピュータサイエンス研究所（2018）「人工知能（AI）が運用に与える影響についての調査研究業務報告書（要約版）」年金積立金管理運用独立行政法人からの委託研究，同法人ホームページ

崔盛旭（2017）「TRMIデータを使ったレジームの分類」京都大学経営管理大学院

陳紀洋，孫華君（2016）「テキスト情報を利用した金利期間構造のモデル化」京都大学経営管理大学院

年金積立金管理運用独立行政法人（2017）「ESG指数選定結果について」同法人ホームページ

野村総合研究所（2016）「ICTの進化が雇用と働き方に及ぼす影響に関する調査研究」総務省からの委託

宮崎浩一（2003）「AHPを利用した最適な債券ポートフォリオの構築に向けて―「決める」から「定める」への橋渡し」日本オペレーションズ・リサーチ学会，10月号，pp. 767-777

山本裕樹，松尾豊（2016）「景気ウォッチャー調査の深層学習を用いた金融レポートの指数化」，The 30th Annual Conference of the Japanese Society for Artificial Intelligence

Bolster, P.J., Janjigian, V., Trahan, E.A. (1995), "Determining Investor Suitability Using the Analytic Hierarchy Process", Financial

Analysts Journal, July-August, pp. 63-75

Campbell, R.H., Liu, Y., Zhu, H. (2016), "… and the Cross-Section of Expected Returns", The Review of Financial Studies, Volume 29, Issue 1, 1 January 2016, Pages 5-68

Chen, K.Y., Liu, S.H., Chen, B., Wang, H.M., Jan, E.E., Hsu, W.L., and Chen, H.H. (2015), "Extractive Broadcast News Summarization Leveraging Recurrent Neural Network Lang uage Modeling Techniques", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 23, pp. 1322-1334

Fama, E., French, K. (1993), "Common risk factors in the returns on stocks and bonds", Journal of Financial Economics 33, pp. 3-56.

Fama, E., French, K. (2015), "A Five-Factor Asset Pricing Model", Journal of Financial Economics, Vol. 116, No. 1, pp. 1-22

Heston, L., S., and Sinha, N.R. (2014), "News versus Sentiment: Comparing Textual Processing Approaches for Predicting Stock Returns", Robert H. Smith School Research Paper

Le, S.V. (2008), "Asset Allocation: An Application Of The Analytic Hierarchy Process", Journal of Business & Economics Research, September Volume 6, Number 9, pp. 87-94

Merton, R.C., Bodie, Z. (1995), "The Global financial system: a functional perspective", "A Conceptual Framework for Analyzing the Financial Environment", edited by Crane, D. B., , Harvard Business Press

Saaty, T.L. (1980), "The Analytic Hierarchy Process", McGraw-Hill, New York

Saaty, T.L., Rogers, P.C., Pell, R. (1980), "Portfolio selection through hierarchies", The Journal of Portfolio Management, Spring, pp. 16-21

Samek, W., Wiegand, T., Muller, K.R. (2017), "Explainable Artificial Intelligence: Understanding,Visualizing and Interpreting Deep Learning Models", Cornel University Library

Tetlock, P.C. (2007), "Giving Content to Investor Sentiment: The Role of Media in the Stock Market", The Journal of Finance, Vol. LXII, NO. 3, JUNE