

集団化する人工知能 鈴木教授による解説シリーズ ⑩ ～ニュースを読んで投資判断する集合知AI～

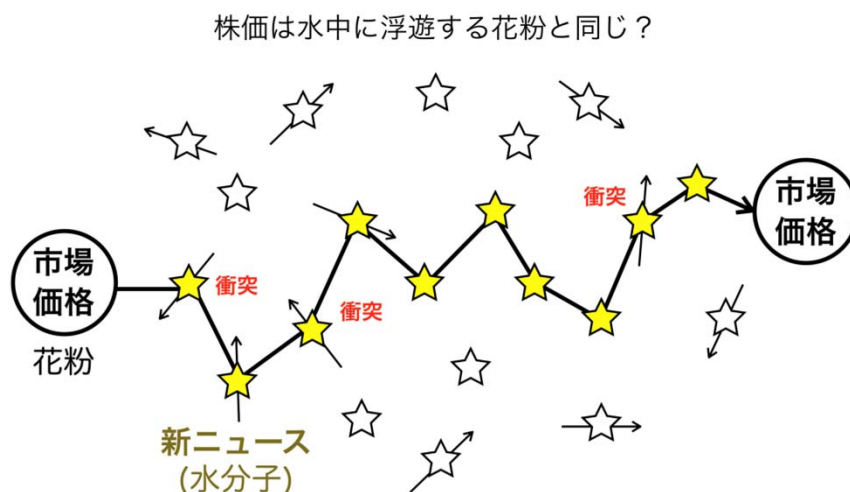
お伝えしたいポイント

2018年12月4日

- ・ ニュースは株価を動かす重要情報
- ・ ニュースを読むAI
- ・ AIの集団でニュースを読み合う
- ・ 集合知AIモデルを高める工夫

<ニュースは株価を動かす重要情報>

伝統的な経済学では、株価はランダムに動くので予測はできないと考えます。なぜなら株価に影響するニュースが発生すると、世界中の誰かがすぐに株を売買して利益を得ようとするからです。このチャンスを活かすには、素早くニュースの意味を解釈し、合理的な売買行動を判断するスピード競争になります。その勝者によってニュースの影響は瞬時に株価に反映されるため、株価は常に適正価格を保つと考えます。これを効率的市場仮説と呼びます。株価を動かす根本的な要因はニュースであり、ニュースの発生は予測できずランダムであるため、株価もランダムに動くと考えます。これをランダムウォーク理論と呼びます。



出所：大和証券投資信託委託

当資料のお取り扱いにおけるご注意

■当資料は、ファンドの状況や関連する情報等をお知らせするために大和投資信託により作成されたものであり、勧誘を目的としたものではありません。■当資料は、各種の信頼できると考えられる情報源から作成していますが、その正確性・完全性が保証されているものではありません。■当資料の中で記載されている内容、数値、図表、意見等は当資料作成時点のものであり、将来の成果を示唆・保証するものではなく、また今後予告なく変更されることがあります。■当資料中における運用実績等は、過去の実績および結果を示したものであり、将来の成果を示唆・保証するものではありません。■当資料の中で個別企業名が記載されている場合、それらはあくまでも参考のために掲載したものであり、各企業の推奨を目的とするものではありません。また、ファンドに今後組み入れることを、示唆・保証するものではありません。

販売会社等についてのお問い合わせ⇒大和投資信託 フリーダイヤル 0120-106212(営業日の9:00～17:00) HP <http://www.daiwa-am.co.jp/>

効率的市場仮説やランダムウォーク理論は、株価の予測可能性を否定しますが、スピード競争なら人間よりも圧倒的にAIが有利です。近年の自然言語処理技術の発達に伴い、AIがニュース（テキストデータ）を分析できるようになりました。ならばAIに大量のニュースを読ませ、高速に投資判断させようとするのは極めて自然です。現在このような試みがアクティブ運用の分野において有望視されています。

<ニュースを読むAI>

簡単なニュースを用いて投資判断するAIをご紹介します。例えば、架空のABC社について以下2つのニュースが流れたとします。

（ニュース1）ABC社、人工知能を活用した個人向けサービスを開始

（ニュース2）ABC社、ビッグデータを活用した法人向けサービスを終了

自然言語処理技術により、文章中の単語の品詞を特定し、単語毎に分解することが可能です。名詞のみに着眼するなら「人工知能」「ビッグデータ」「活用」「個人」「法人」「サービス」「開始」「終了」が抽出されます。

それぞれの単語を含むなら1を、含まないなら0を付与することで、各ニュースをベクトルで表現できます。上記の単語順に従うと、

$$\text{ニュース1} = \{ 1, 0, 1, 1, 0, 1, 1, 0 \}$$
$$\text{ニュース2} = \{ 0, 1, 1, 0, 1, 1, 0, 1 \}$$

となります。

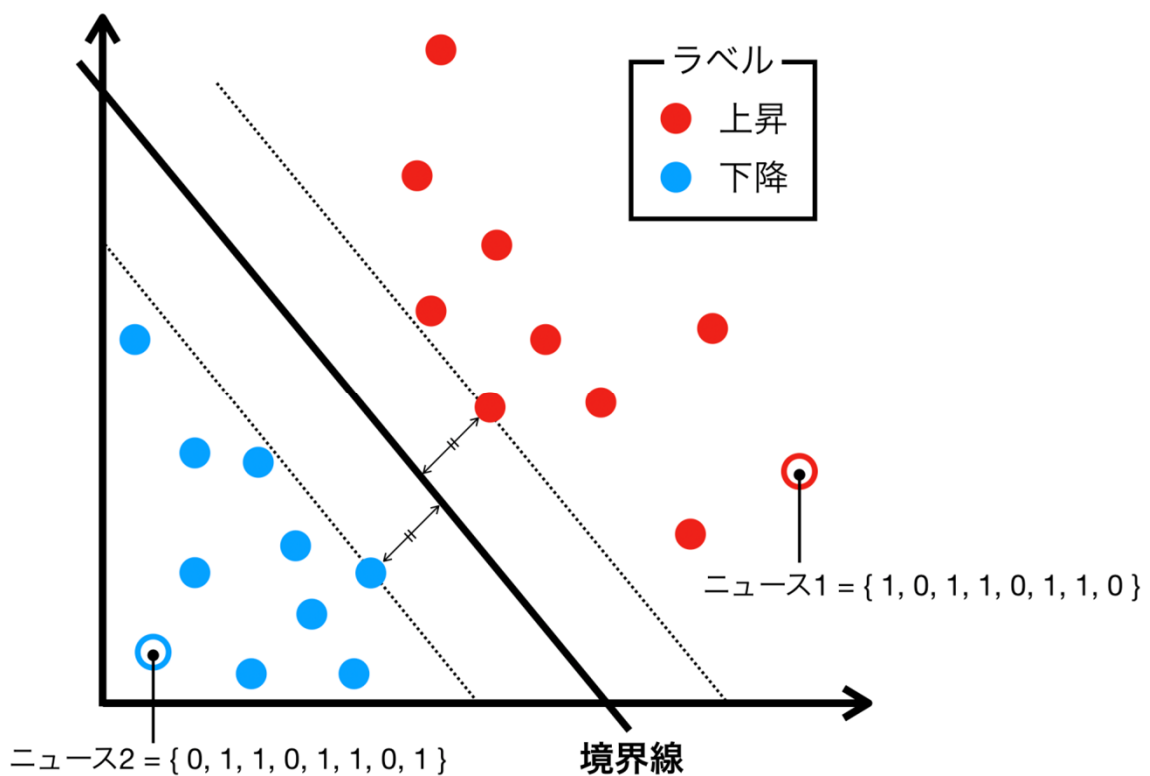
ニュース発生後の株価変動を判別するには、過去の実績を調べるのが良さそうです。しかし同一の単語ペアを含むニュースがたくさん得られれば良いですが、一般的に同一単語ペアのニュースは起こり難いです。その対策として、機械学習の方法が用いられます。

最も単純な方法として「ナイーブベイズ法」が有名です。単語毎に株価への反応を調べ、それらを掛け合わせることでニュース全体の影響力を推定します。単語毎に確率計算できるため、対象単語を増やしても計算負担が軽いという利点があります。しかし単語間の関連性を一切考慮できません。

例えば「人工知能」と「個人」を同時に含む場合、「人工知能」と「法人」を同時に含むよりもニュースの影響力は高いかもしれません。近年はAIスピーカーなど個人向け商品が注目されており、多くの人に恩恵を与える可能性があるからです。

このように単語間の関連性も考慮するなら「サポートベクターマシン」のように空間を分割する機械学習法を用います。

まず先述の方法によって、過去に発生した全ニュースを0と1のベクトルに変換します。次に各ニュース発生後の株価変動を参照し、「上昇」または「下降」のラベルを各ニュースに付与します。下記がその概念図です。本来は高次元の空間ですが、可視化のため2次元で図示しています。



出所：大和証券投資信託委託

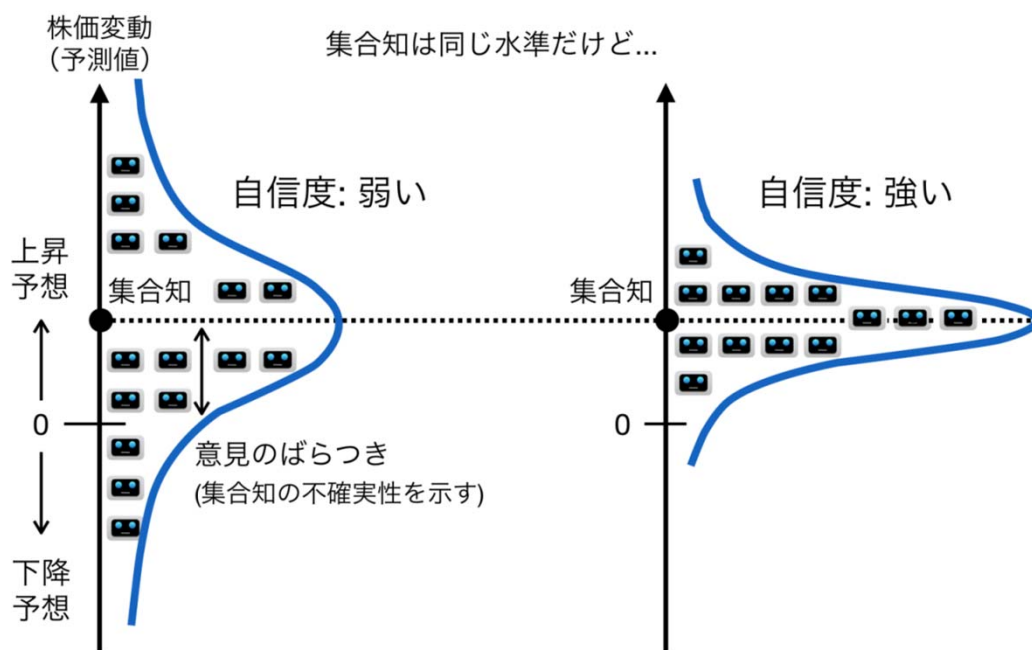
機械学習のアルゴリズムによって、「上昇」と「下降」のラベルを上手く分離できる境界線を学習します。その境界線に基づいて、新規のニュースのラベルを判別します。例えば概念図のように、ニュース1が上昇グループに入ったならば、AIはその後の株価が上昇すると判断します。同様にニュース2の場合は下降すると判断します。

このように境界線に基づいてラベルを判別するため、各ニュースが境界線から離れるほど判別が明確になります。さらに遠ざかる方向も調べることで、ニュースを構成する各単語の極性（株価に対する影響がポジティブかネガティブか）の度合いを定量化できます。

例えば「高い」という単語は、物価に対してはネガティブな印象ですが、株価に対してはポジティブな意味を持ちます。そのためAIに売買判断をさせるには、株価と単語を紐づけることで、株価に特化した極性辞書を作る必要があります。上記の機械学習アルゴリズムは言語を問いませんので、日本語以外の極性辞書も作成可能です。

<AIの集団でニュースを読み合う>

これまでのマーケットレターを通じて、AIを集団化させることで機械知能を強化する工夫をご紹介してきました。特に集団の投票結果（多数決や平均値）は安定して良い答えを導けることは統計学的にも証明されており、一般的に「集合知」として知られています。さらに私が提唱する集合知AIモデルでは「意見のばらつき具合（リスク）」も定量化することで、集合知に対する自信度も投資判断に利用します。これにより自信がない無謀な投資を避けることができます。



出所：大和証券投資信託委託

新しいニュースが出現する度に、たくさんのAIにニュースを読ませ、投資判断してもらいます。最終的な売買判断にはこれらAIの集合知を用いますが、次の2点を検討する必要があります。

- (1) 集合知を発揮するには、これらAIの独立性を高める必要がある。
- (2) 集合知の不確実性に応じて、投資を行う基準を定める必要がある。

<集合知AIモデルを高める工夫>

まず (1) AIの独立性を高める工夫として、次の3要素が考えられます。

1. 学習データの多様性

全ニュースから、ラベルを保持したまま、ニュースをランダムに復元抽出します。学習データを構成するニュースの組合せが変わるため、多様な学習データを生成できます。一般に「バギング」と呼ばれる手法です。

2. 学習モデルの多様性

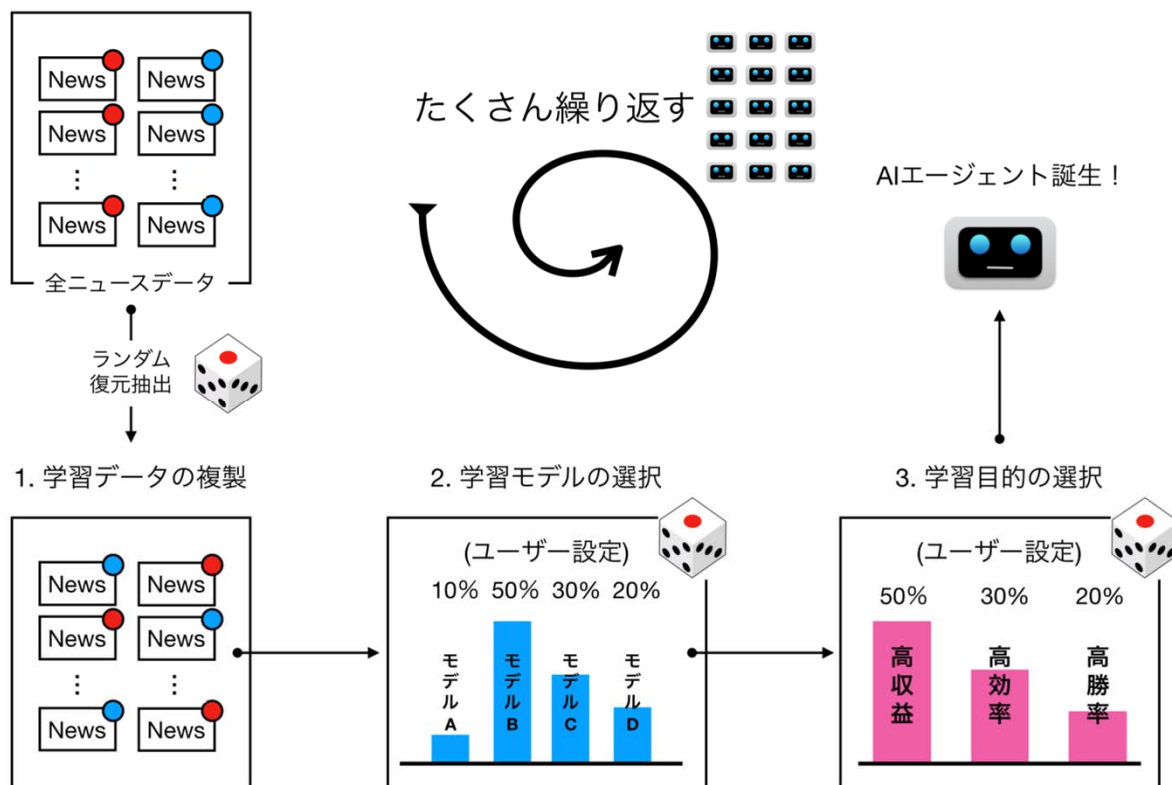
それぞれのAIを構成する機械学習モデルを1つに定めず、多様性を持たせます。例えば前述のサポートベクターマシンやナイーブベイズ法のみならず、ニューラルネットワークや決定木など、概念が異なるモデルを確率的に選択します。この選択割合はユーザーが柔軟に調整できます。

※1ページ目の「当資料のお取り扱いにおけるご注意」をよくお読みください。

3. 学習目的の多様性

アクティブな運用成績の評価基準として、高収益・高効率・高勝率など多様に存在するため、AIEージェントの学習目的も確率的に選択します。この選択割合もユーザーが柔軟に調整でき、2.と3.を通じてAIEージェント1体の仕様が決まります。その後、この学習目的が最大になるように、2.で選ばれた学習モデルのハイパーパラメータ※を最適化します。最後に、1.で複製された学習データを上手く予測できるようにモデルパラメータを学習すると、AIEエージェントが完成します。

※ 機械学習には、教師ラベルとの整合性によって自動的に決定されるモデルパラメータの他に、機械学習の目的に応じてモデルの構造をデザインするためのハイパーパラメータが存在します。ハイパーパラメータがモデルパラメータを支配するため、先に最適化しておく必要があります。



※1ページ目の「当資料のお取り扱いにおけるご注意」をよくお読みください。

これら1.～3.をたくさん繰り返すことで、仕様の異なるAIエージェントを多様に生成できるため、精度の高い集合知を獲得できると考えられます。

次に（2）投資の判断基準について検討します。集合知の不確実性を考慮して、以下の4種が考えられます。いずれもユーザーが設定した閾値を超えたら投資を行い、集合知の符号に基づいて売買選択します（プラスなら買い、マイナスなら売り）。

基準1. マハラビス距離 = (|集合知| - |要求値|) ÷ 標準偏差

要求値は、売買手数料など最低限獲得したい利益を設定します。集合知はAI集団の平均値とし、マハラビス距離が大きいほど集合知に自信があるとみなします。

（参考イメージを次ページに記載しています。下記の基準も同様）

基準2. 基準1の分子のみ = |集合知| - |要求値|

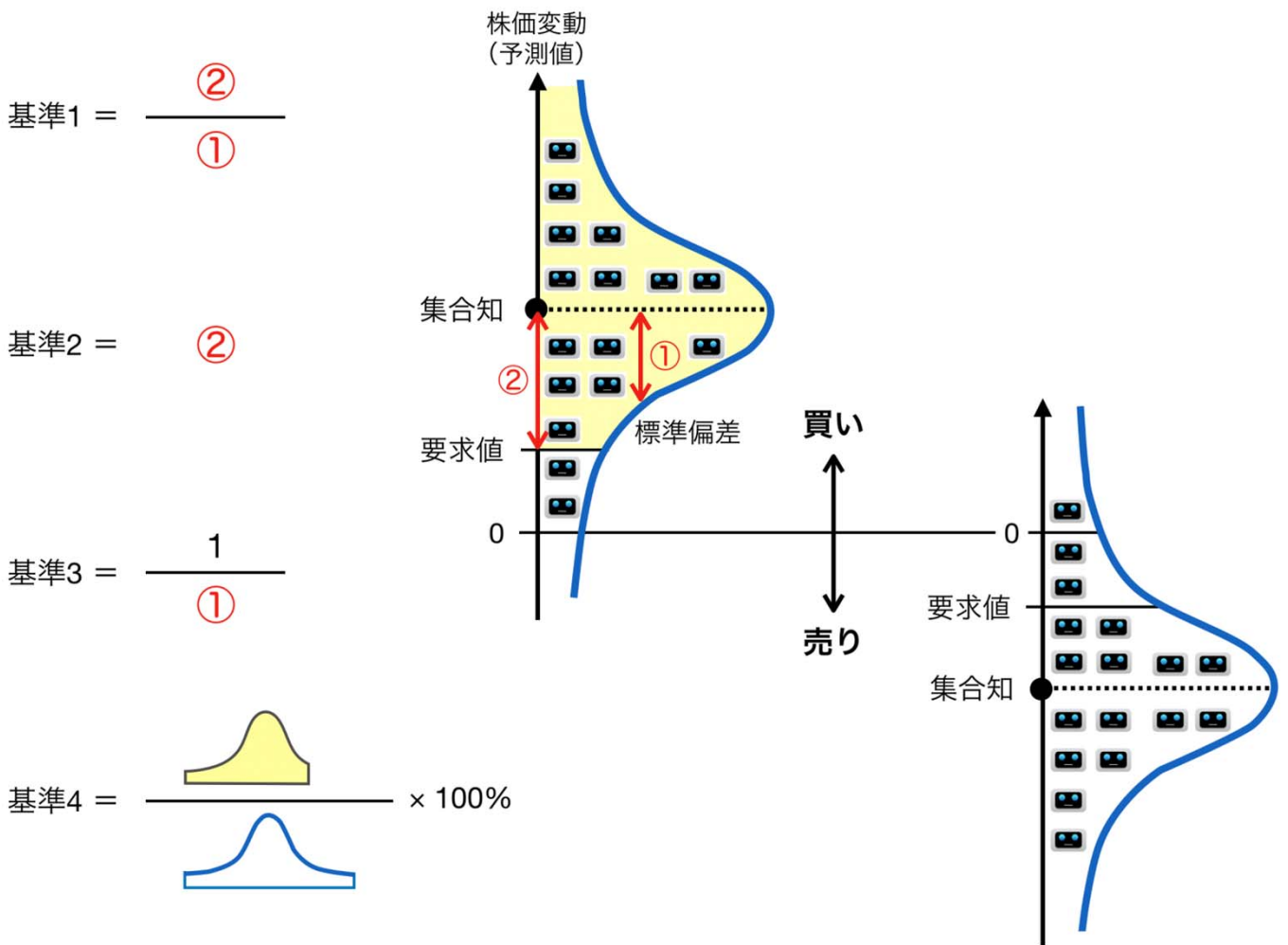
標準偏差を考慮せず、集合知が要求値から乖離するほど自信があるとみなします。

基準3. 基準1の分母のみ = 1 ÷ 標準偏差

集合知の大きさを考慮せず、標準偏差が小さいほど集合知に自信があるとみなします。ただし集合知が要求値を超えなければ投資しません。

基準4. 多数決の一致率 = 要求値以上の予測値を示すAIの割合

要求値以上の利益が得られるか否かの2択問題とし、利益が得られると回答したAIが多いほど集合知に自信があるとみなします。



出所：大和証券投資信託委託

※1ページ目の「当資料のお取り扱いにおけるご注意」をよくお読みください。

基準1が集合知の大きさと標準偏差を同時に考慮するため、最も丁寧な投資判断となります。しかし株価予測はそもそも難しいので、基準4などの大雑把な方法が良い可能性もあります。いずれにせよ、AIの集合知に自信が得られない要因として、

- ・常に株価に影響するニュースが出現するとは限らない
- ・出現したニュースが特殊で、学習データに同様な事例が少ない

ことが挙げられます。このような場合でもAIは取りあえずの答えを返すため、AIを常に信用するのは大変危険です※。そこで集合知AIモデルのように意見の不一致も考慮する必要があります。これはお金に関わる資産運用のみならず、人命に関わる医療診断においても重要なポイントだと考えています。

※ 人間は過去の事例が無くても演繹的に物事を推論できます。全てをAIに任せるのではなく、協働によって互いの長所を活かし短所を補うことで、AI技術の恩恵を享受できると思います。

【鈴木教授 プロフィール】

鈴木智也(すずきともや)

新潟県新潟市生まれ。IFTA国際検定テクニカルアナリスト(MFTA)。平成17年東京理科大学大学院理学研究科物理学専攻博士課程修了。理学博士。同年東京電機大学工学部電子工学科助手、平成18年より同志社大学工学部情報システムデザイン学科専任講師、平成21年より茨城大学工学部知能システム工学科准教授を経て、平成28年より同大学教授、さらに平成29年より大和証券投資信託委託(株)クウォンツ運用部特任主席研究員を兼務。平成30年より茨城大学大学院理工学研究科機械システム工学専攻長および領域長、CollabWiz株式会社代表取締役。

研究分野は、時系列解析、テキスト解析、機械学習、人工知能、金融工学など実践的なデータサイエンスに従事。電子情報通信学会、情報処理学会、人工知能学会、日本テクニカルアナリスト協会、日本証券アナリスト協会各会員。

【Market Letter 鈴木教授による解説シリーズ バックナンバー】

第1回 AI運用に挑む	http://www.daiwa-am.co.jp/market/html_ml/ML20171207_1.html
第2回 集団化する人工知能	http://www.daiwa-am.co.jp/market/html_ml/ML20180125_1.html
第3回「2年目のジンクス」を集合知AIで緩和	http://www.daiwa-am.co.jp/market/html_ml/ML20180301_1.html
第4回 時系列データの見えない法則をつかむ	http://www.daiwa-am.co.jp/market/html_ml/ML20180409_1.html
第5回 愚かな人間心理・カモにするAI	http://www.daiwa-am.co.jp/market/html_ml/ML20180501_2.html
第6回 ナイトメア★アノマリーを狙え	http://www.daiwa-am.co.jp/market/html_ml/ML20180601_1.html
第7回 ブルーオーシャンAI戦略	http://www.daiwa-am.co.jp/market/html_ml/ML20180703_1.html
第8回 深層学習による株価予測 (前編)	http://www.daiwa-am.co.jp/market/html_ml/ML20180802_2.html
第9回 深層学習による株価予測 (後編)	http://www.daiwa-am.co.jp/market/html_ml/ML20181001_1.html

※1ページ目の「当資料のお取り扱いにおけるご注意」をよくお読みください。