【鈴木教授による解説シリーズ12】集団化する人工知能

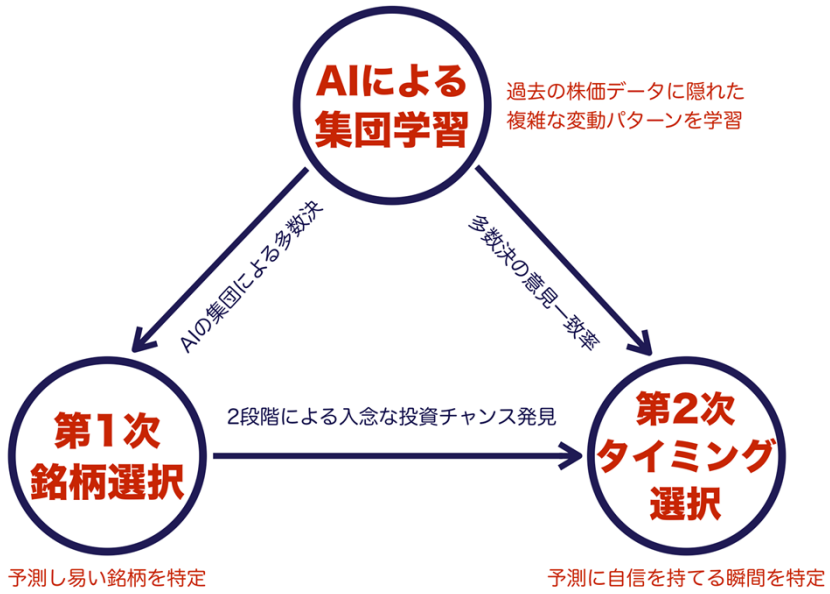
集合知AIモデルのシミュレーション(前編)

2019年 8月 5日　　**【図解シリーズ】投資理論とコンピュータの歴史** の続きは10月を予定しています。

お伝えしたいポイント

- 2回に分けて、集合知AIモデルが機能するメカニズムを解説します。
- 今回は、集合知 (集団学習) のメカニズムを解説します。
- 第1次銘柄選択 (集合知で予測しやすい銘柄を特定) の効果を解説します。
- 次回も続きます (第2次タイミング選択など)。

本解説シリーズ第1回（AI運用に挑む）にて「集合知AIモデル」をご紹介しました。その原著論文「Consensus Ratio and Two-steps Selection to Detect Profitable Stocks: Modern Technical Analysis Using Machine Learning Approach」では、翌日株価の変動方向の予測において約60%～80%の正答率を得ています (原著論文のTable5)。そこで2回に分けて、何故このような高い正答率を実現できるのか、そのメカニズムについて考察したいと思います。なお原著論文は、国際テクニカルアナリスト連盟 (IFTA) ホームページ (https://ifta.org/public/files/journal/d_ifta_journal_18.pdf) より無料で入手できます.



（出所）大和証券投資信託委託

注：
左図は、集合知AIモデルのコンセプト:
解説シリーズ第1回（AI運用に挑む）
https://www.daiwa-am.co.jp/specialreport/quants/2017/quantst_20171207.html
を一部変更して再掲しています。

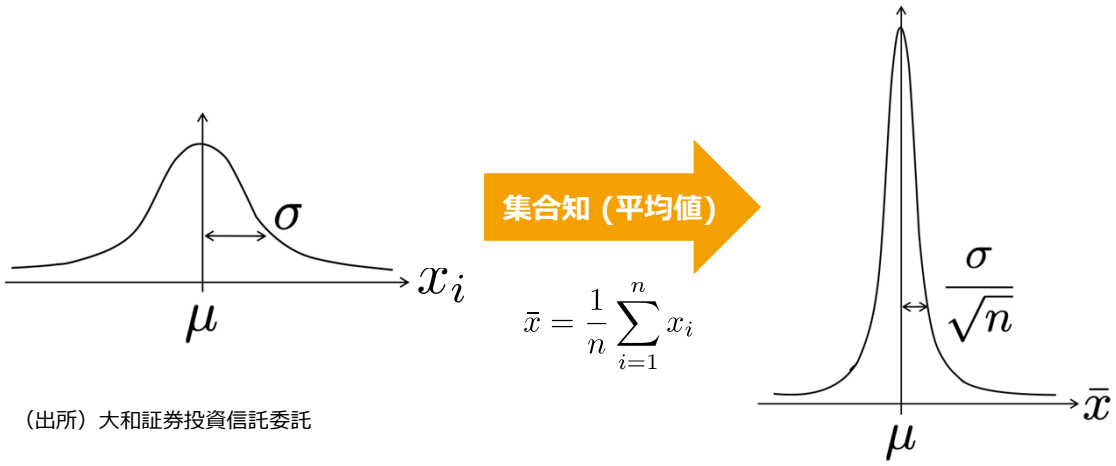
集合知のメカニズム

回帰問題 (連続値の予測) の場合

3人寄れば文殊の知恵のように、複数人の判断を統合すると判断能力が向上します。これは「集合知」という統計的な現象であり、直感的に言えば、各自の予測誤差が互いに打ち消しあって縮小する現象です。数式で表現するならば、

$$\text{各自の予測値 } x_i = \text{正解値 } \mu + \text{予測誤差 } \epsilon_i$$

とし、予測誤差ϵ_iは平均0かつ標準偏差σの乱数 (ランダムな値) とします。なお、予測する人数はn人とします ($i = 1 \sim n$)。もし各自の予測が互いに独立に行われ、各々の予測値を平均することで集合知を得るならば、集合知による予測値$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$は中心極限定理により、平均$\mu$かつ標準偏差$\frac{\sigma}{\sqrt{n}}$の正規分布に従います。予測する人数$n$が増えるほど標準偏差が小さくなるため、集合知による予測値$\bar{x}$は正解値μに近づきます。

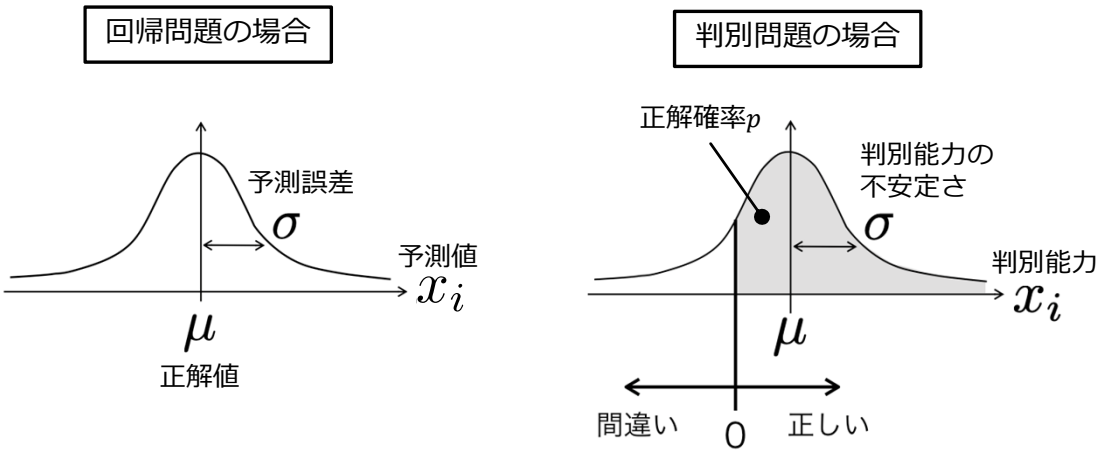


（出所）大和証券投資信託委託

一方、各自が相談し合って予測するなど互いに独立ではないならば、実質的に独立な予測者の人数nが減ると考えられるため、集合知の標準偏差$\frac{\sigma}{\sqrt{n}}$は縮小しにくくなります。したがって集合知の効果は低下します。

判別問題 (離散値の予測) の場合

前ページでは回帰問題 (予測値x_iが連続値) の観点から集合知の効果を示しました。しかしここで示す集合知AIモデルでは株価の変動方向 (上昇 or 下落) を予測対象とするため「2クラス判別問題」に相当します。そこでx_iを各自の予測値ではなく、各自の判別能力とみなします。そしてx_iが0を超えたら判別問題に正解できるとします。したがって正解確率pは$x_i > 0$の面積となります。なおx_iは平均μかつ標準偏差σの乱数とし、平均μは基本的な判別能力であり、標準偏差σはその不安定さを意味します。つまりμが大きくσが小さいほど優れた判別であるため、正解確率pが高くなります。



（出所）大和証券投資信託委託

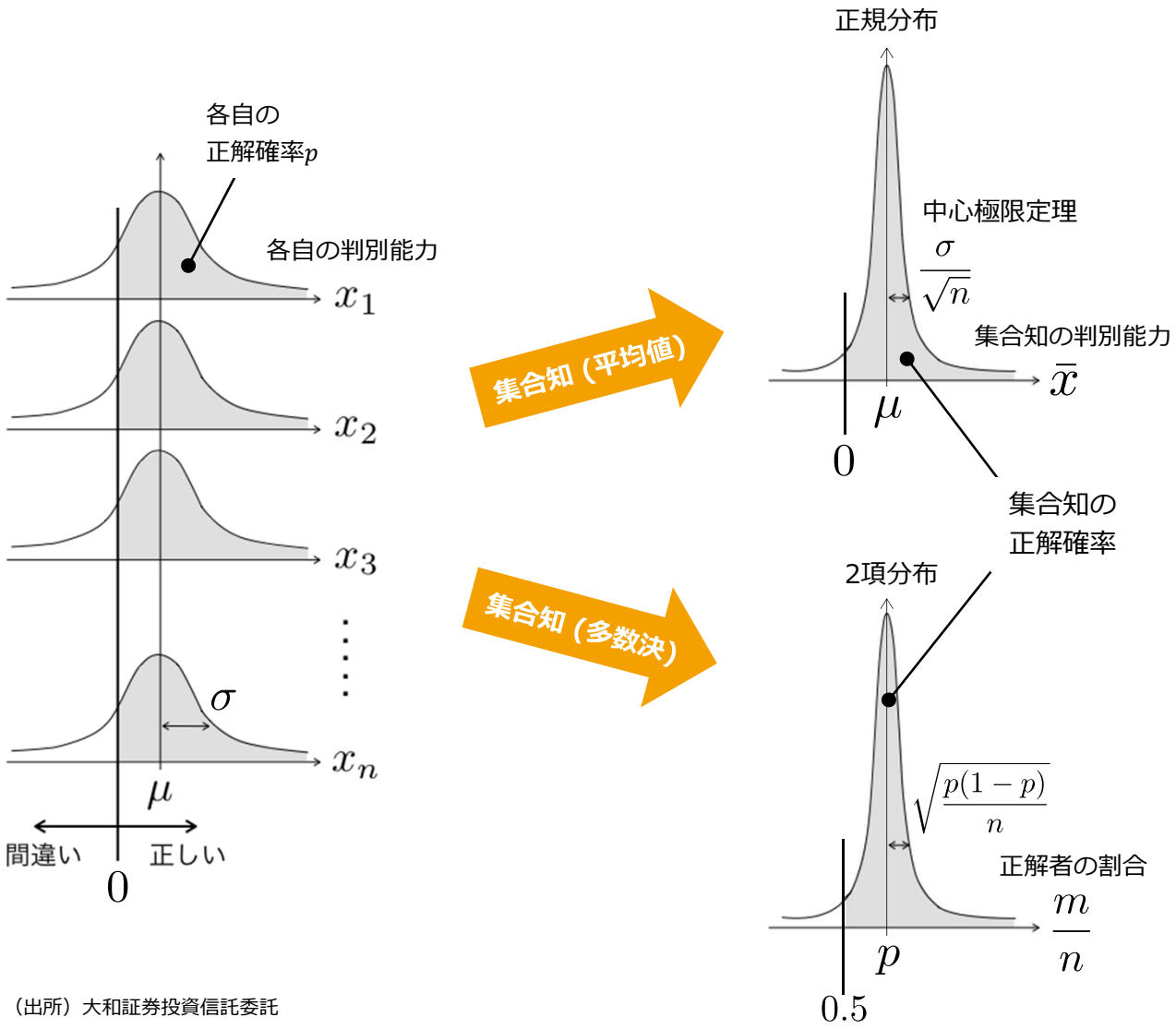
さて、集合知による判別を考えてみましょう。平均値$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$によって判別する場合と、多数決によって判別する場合の2パターンが考えられます。それぞれの判別結果が正解となる条件は以下のとおりです。

■ 平均値による判別： $\bar{x} > 0$ ならば集合知は正しい。

回帰問題と同様に、中心極限定理によって集合知の効果を説明できます。

■ 多数決による判別： 正解者 ($x_i > 0$) の人数をmとすると、 $\frac{m}{n} > 0.5$ならば集合知は正しい。

n人中m人が正解する確率は二項分布に従うため、二項分布の観点から集合知の効果を説明できます。nの増加と共に標準偏差$\sqrt{\frac{p(1-p)}{n}}$は縮小するため、各自の正解確率が$p > 0.5$ならば集合知の正解確率は100%に近づきます。



（出所）大和証券投資信託委託

(※) 二項分布の横軸は離散値です

回帰問題と同様に、各自の判断が独立ではないならば、実質的に独立な多数決の人数nが減ると考えられるため、集合知の標準偏差$\frac{\sigma}{\sqrt{n}}$および$\sqrt{\frac{p(1-p)}{n}}$は縮小しにくくなります。したがって集合知の効果を得るには、各自の判断を可能な限り独立にする工夫が必要になります。

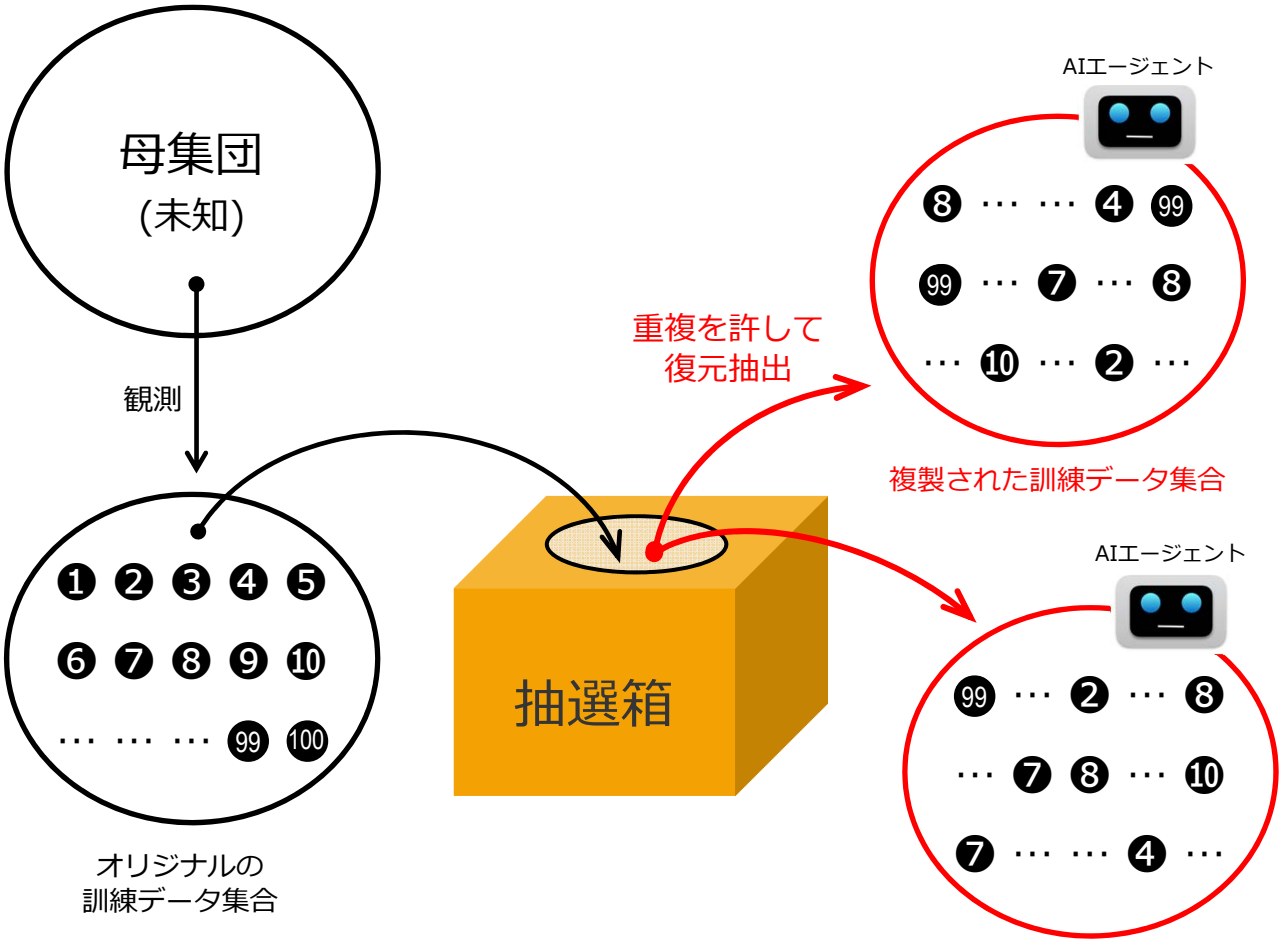
独立に判断するAIを構築する工夫 (集団学習)

集合知AIモデルは、多数の人工知能（AI）を構築することで集合知の効果を発揮します。しかしAIを鍛える訓練データは基本的に1セットしかありませんので、無策であれば全て同じAIが構築されてしまいかねません。この場合、n=1と等価であるため、集合知の効果を発揮できません。

そこで、できるだけ特徴が異なるAIを構築する工夫として「バギング」という集団学習法をご紹介します。集団学習は本解説シリーズ第10回（ニュースを読んで投資判断する集合知AI）でも登場しました。

まずオリジナルの訓練データ集合すべてを抽選箱に入れ、同一のデータ数になるまでランダムにデータを復元抽出することで異なる訓練データ集合を複製します。ここでデータ数はオリジナルと同一であるため、一部のデータは同一集合内で重複します。この複製を繰り返したのち、複製されたデータ集合毎でAIを訓練すれば、特徴の異なる（独立性の高い）AIを構築できる可能性があります。つまり復元抽出に伴うランダム性を利用して独立性を高める作戦ですが、オリジナルの訓練データ集合は共通であるため、復元されたデータ集合間においても重複が発生します（模式図では2番・4番・8番・7番・10番・99番が重複)。この集合間での重複がAIの独立性を下げる要因となるため、たとえn体のAIを構築したとしても、実質的に独立な個体数はn未満になります。

なおオリジナルの訓練データ集合は、未知の母集団に含まれる一部の標本に過ぎません。そこでランダムな復元抽出により多様性を回復することで、未観測の母集団を再現したいという思惑もあります。これをブートストラップ法と呼び、バギング (Bootstrap Aggregatingの略) の由来になっています。



（出所）大和証券投資信託委託

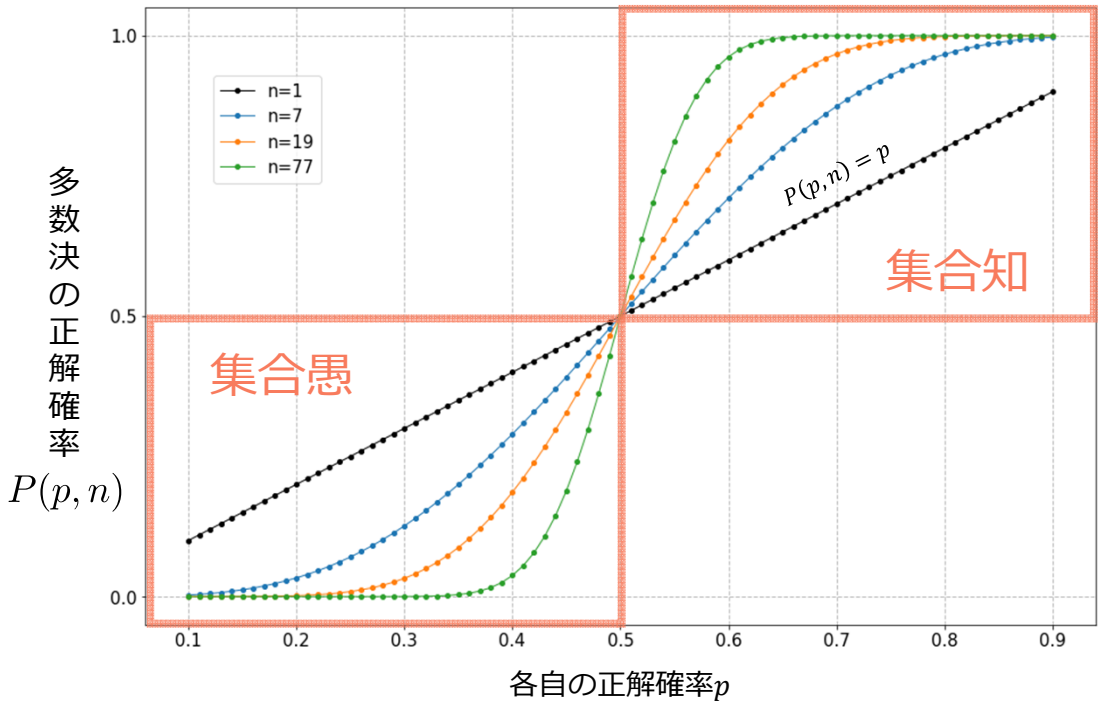
集合知AIモデルへの応用

集合知と集合愚

先述のように、集合知AIモデルは翌日株価の変動方向 (上昇 or 下落) を予測対象とするため「2クラス判別問題」に相当します。なお実際には各AIの判別能力x_iを計測できないため、平均値ではなく多数決によって集合知を得るとします。この場合、n体のAIによる集合知 (多数決) の正解確率は

$$P(p,n) = \sum_{m>0.5n}^{n} {}_nC_m p^m (1-p)^{n-m}$$

となります。ここでpは各自の正解確率、mは正解したAIの数を意味します。この確率Pはpとnのみに依存するため、両変数を変えながら$P(p,n)$のグラフを作成しました。



（出所）大和証券投資信託委託

上図において、確かに集合知の効果を確認できます。$p > 0.5$ ならば、$P(p,n) > p$となります。また、nが大きいほど$P(p,n)$が高まる結果は先述の説明と整合的です。一方、$p < 0.5$ ならば、$P(p,n) < p$となりますが、これはそもそも正解できないタスク ($p < 0.5$) です。AIは過去の経験に基づいて訓練しますので、過去の経験に逆らう現象については反対に予測します。このように$p < 0.5$であれば、多数決によってますます正解確率は低下するため、この場合は「集合愚」と呼ぶことにします。

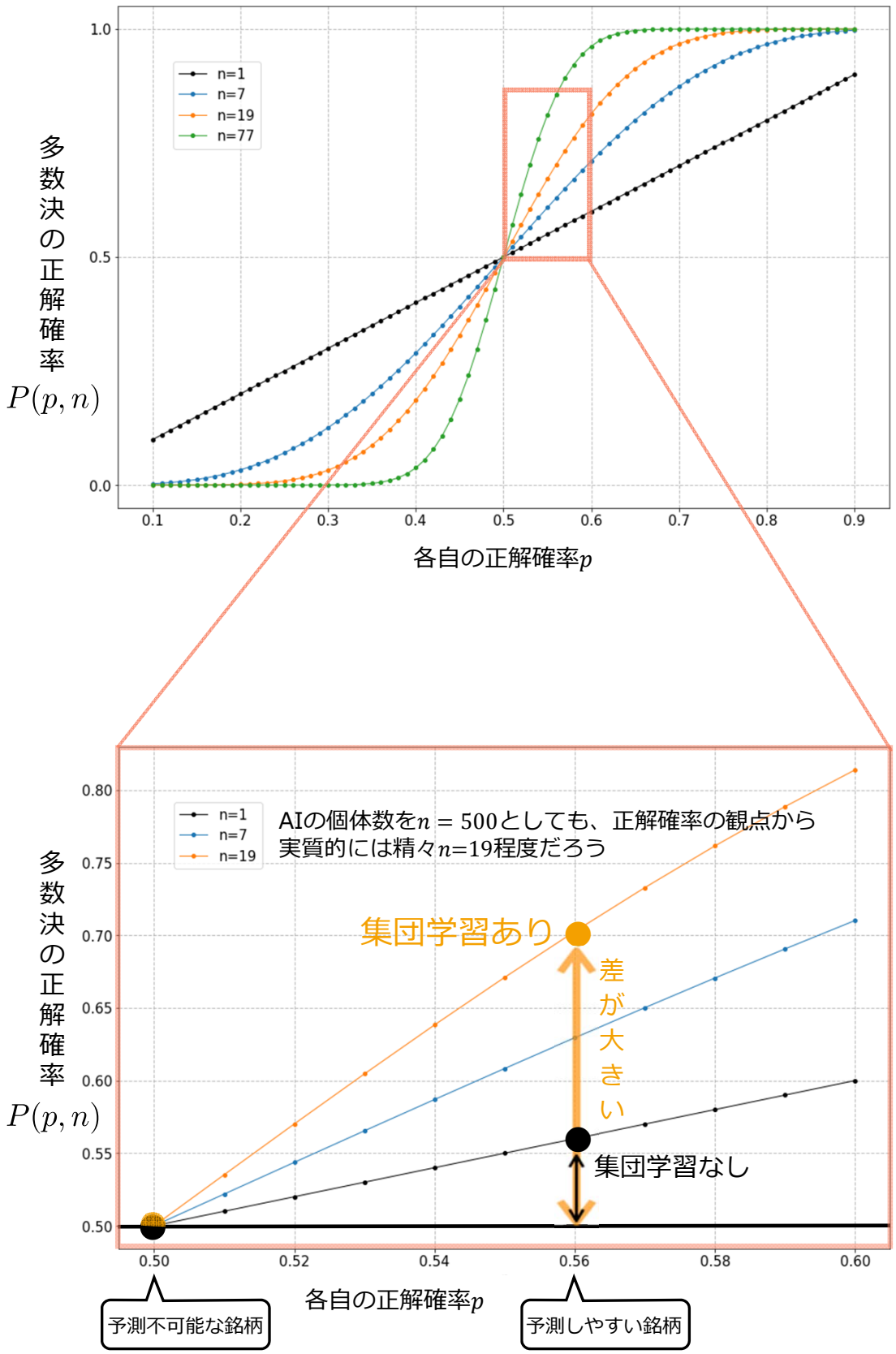
集合知によって予測しやすい銘柄を特定する (第1次銘柄選択)

次に、実際の株価予測らしい正解確率として、$p = 0.5$~0.6の範囲に着目します。株価変動が完全にランダムならば$p = 0.5$でありますが、少しでも法則性があるならば$p > 0.5$となります。前者は予測不可能な銘柄であり、後者は予測しやすい銘柄と言えます。もし両者の銘柄が混在するならば、後者のみを予測対象にするのが得策です。そこで後者を特定するために、集合知を応用します。

一般に、集合知 (集団学習を含む) は判断能力を向上させる方法として知られていますが、難しい問題と易しい問題を分離する方法としても利用できます。次の模式図は、予測不可能な銘柄 $(p = 0.5)$ と予測しやすい銘柄 $(p = 0.56)$ を分離する例ですが、集団学習なし $(n = 1)$ に比べて、集団学習を適用した方が両者の正解確率$P(p, n)$の差を拡大できます。株価銘柄毎の予測可能性は時間変化によって揺らぐならば、正解確率$P(p, n)$の差は曖昧になりますので、集団学習によってこの差を拡大できる点は銘柄選択において大きなメリットです。そこで集合知AIモデルでは、これを「第1次銘柄選択」と呼び、集団学習による正解実績が高い銘柄のみを運用の対象にします。

なお原著論文では、500体のAIによる集団学習を実施していますが、先述の訓練データが重複する問題や多数決によって実現できる正答率の観点から、実質的に独立に機能するAIは19体程度 $(n = 19)$ だと考えられます。しかしそれでも正解確率を向上でき、さらに銘柄選択においても集団学習を適用するメリットがあります。

次回は、簡単なエージェントシミュレーションによって、集合知AIモデルの妥当性を検証します。本当に第1次銘柄選択によって予測しやすい銘柄を特定できるのか検証し、第2次タイミング選択のメカニズムについても解説します。集合知AIモデルの真価は「選択力」にあります。そして原著論文で得られた約60%～80%の正答率をエージェントシミュレーションによって再現します。



（出所）大和証券投資信託委託

鈴木教授　プロフィール

鈴木智也（すずきともや）

茨城大学大学院理工学研究科機械システム工学領域教授．理学博士．IFTA国際検定テクニカルアナリスト．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，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，平成30年よりCollabWiz(株)代表取締役を兼務．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，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．

バックナンバー

鈴木教授による解説シリーズ

- **001「AI運用に挑む」**
  https://www.daiwa-am.co.jp/specialreport/quants/2017/quantst_20171207.html
- **002「集団化する人工知能」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20180125.html
- **003「２年目のジンクスを集合知AIで緩和」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20180301.html
- **004「 時系列データの見えない法則をつかむ」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20180409.html
- **005「愚かな人間心理・カモにするAI」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20180501.html
- **006「ナイトメア★アノマリーを狙え」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20180601.html
- **007「ブルーオーシャンAI戦略」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20180703.html
- **008「深層学習による株価予測 (前編)」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20180802.html
- **009「深層学習による株価予測 (後編)」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/quantst_20181001.html
- **010「ニュースを読んで投資判断する集合知AI」**
  https://www.daiwa-am.co.jp/specialreport/quants/2018/1204.html
- **011「投資理論とコンピュータの歴史」**
  https://www.daiwa-am.co.jp/specialreport/quants/20190701_01.html

当資料のお取扱いにおけるご注意

■当資料は投資判断の参考となる情報提供を目的として大和投資信託が作成したものであり、勧誘を目的としたものではありません。投資信託のお申込みにあたっては、販売会社よりお渡しする「投資信託説明書(交付目論見書)」の内容を必ずご確認のうえ、ご自身でご判断ください。

■当資料は信頼できると考えられる情報源から作成しておりますが、その正確性・完全性を保証するものではありません。運用実績などの記載内容は過去の実績であり、将来の成果を示唆・保証するものではありません。記載内容は資料作成時点のものであり、予告なく変更されることがあります。また、記載する指数・統計資料等の知的所有権、その他一切の権利はその発行者および許諾者に帰属します。