

オルタナティブデータと機械学習の活用状況

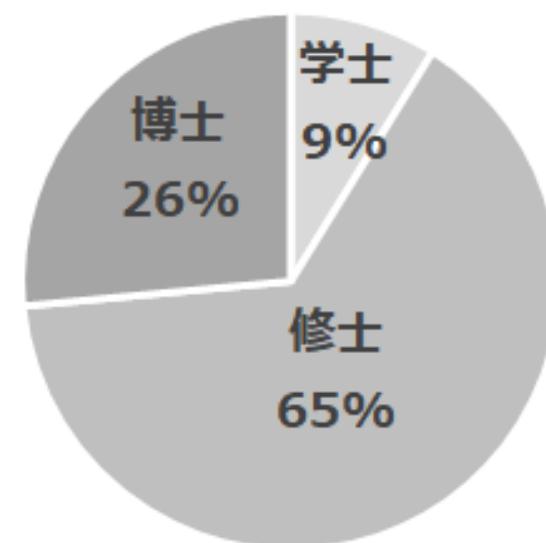
MTECでは・・・

株式会社 三菱UFJトラスト投資工学研究所（MTEC）
久木田 徹

MTECの概要

社名	株式会社 三菱UFJトラスト投資工学研究所 Mitsubishi UFJ Trust Investment Technology Institute Co., Ltd.
本社	〒107-0052 東京都港区赤坂4-2-6 住友不動産新赤坂ビル10階
代表者	取締役社長 成川 順一 取締役副社長 久保田 尚
設立	1988年1月14日
株主	三菱UFJ信託銀行グループ 100%
資本金	4億8,000万円
従業員数	48名（2020年4月1日現在）
売上高	13億2,300万円（2020年3月期）
主要取引先	三菱UFJ信託銀行、三菱UFJ国際投信、三菱UFJ銀行
沿革	・1988年1月 株式会社エムティー ビーインベストメント テクノロジー研究所を設立 ※三菱信託銀行（現三菱UFJ信託銀行）100%出資 ・2005年10月 親会社合併に伴い、株式会社三菱UFJトラスト投資工学研究所に社名変更 ※三菱UFJ信託銀行100%出資

プロパー研究員・FEの内訳



- ・物理の博士が多い
 - ・3名博士課程在学中
 - ・1名Stanford客員研究員
 - ・中途採用は公認会計士やITエンジニアなど
- （2020年4月1日現在）

機械学習・オルタナティブデータの利用例:実ファンドへの適用

実際に、機械学習 & オルタナティブデータをファンドの運用に活用し始めたのは2015年くらいから

Bloombergの記事 (<http://www.bloomberg.co.jp/news/123-NJ8A1V6KLVSD01.html>) より

関学教授ら三菱UFJ信託のファンド助言、ビッグデータで行動分析

2015/02/04 13:31 JST

記事を印刷する 記事をメールで送信

シェア Twitter チェック Google Share 共有/ブックマーク

(ブルームバーグ): 関西学院大学の岡田克彦教授らが設立したマグネマックスキャピタルマネジメントは13日から、三菱UFJ信託銀行の日本株 運用ファンドへの投資助言を開始する。通信社などが配信するニュースを蓄積したビッグデータを基に、コンピュータを使って市場参加者の心理の揺れを推定し、投資アドバイスする。

岡田最高経営責任者(CEO)によると、助言するのは「ビッグデータ活用型ファンド(仮称)」。三菱UFJ信託銀行が約10億円を拠出してテスト運用し、結果を確認した後に年金基金や投資信託の運用商品に展開するという。最大運用可能額は400億円。

投資モデルは、同大学院の岡田研究室が開発したニュースなどに使われた単語を集計分析し点数化した理論モデルを、三菱UFJトラスト投資工学研究所とともに精緻化し、投資戦略を増やした。投資銘柄は平均100銘柄、目標収益率については非公表としている。

マグネマックスは助言業務の拡大を検討しており、岡田氏は「2、3年のうちに1000億円規模の助言を目指す」と述べた。

日本株式についてのロング・ショートファンド

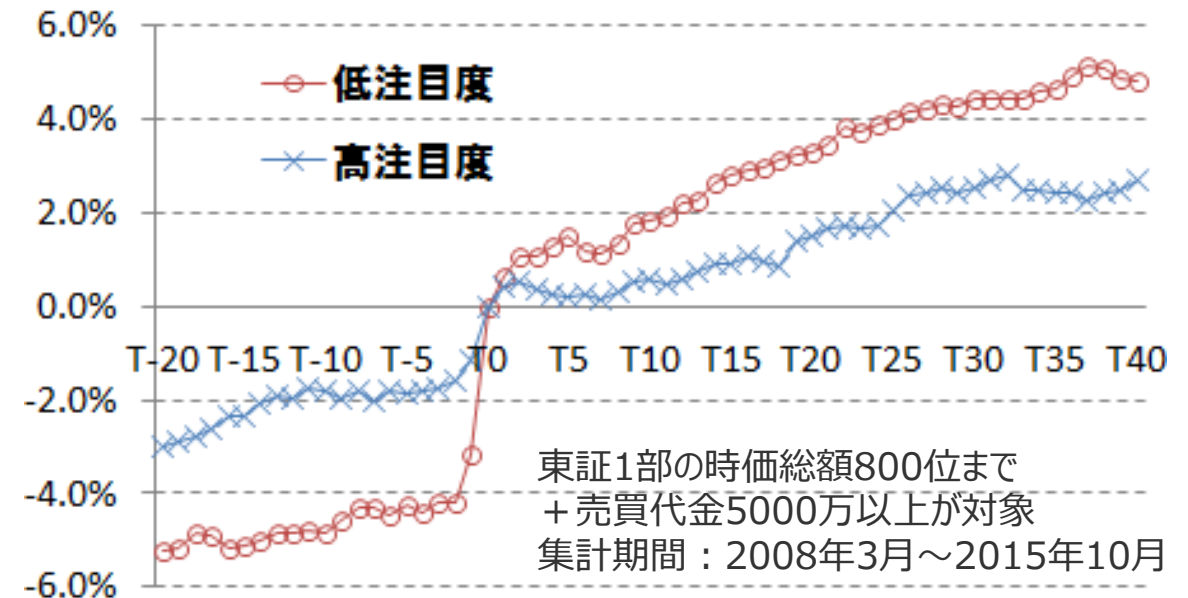
□現物戦略

ファンダメンタル+ニュースのセンチメントや注目度

□先物戦略

Deep Learning + Random Forest + α

テキストを用いた注目度選別の効果

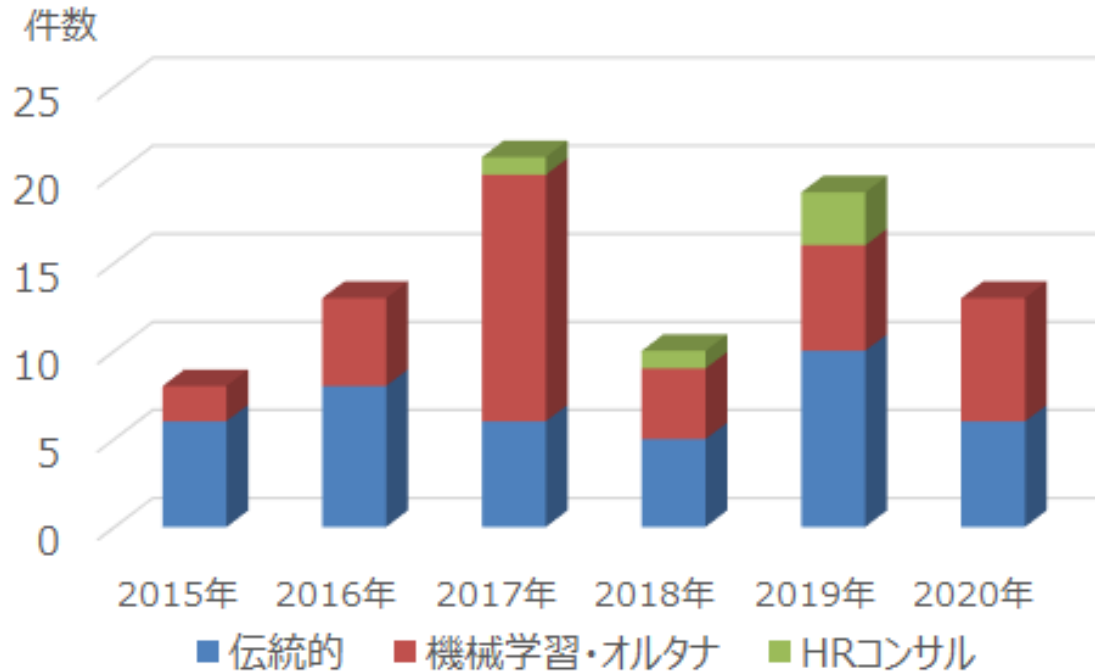


横軸：経過日数 (T-1: 発表日/T0: 購入日) 縦軸：TOPIX超過リターン

機械学習・オルタナティブデータの利用例: 対外活動

日本株運用以外にも機械学習 & オルタナティブデータを業務に活用すべく、研究テーマを拡大中

年別対外発表件数



ざっくりとした分類

□ 伝統的

アセットプライシング
数理ファイナンス

□ 機械学習・オルタナティブデータ

自然言語処理（ニュース、SNS、Q&A、・・・）
ネットワーク（企業、プロダクト、イベント、・・・）
Tick Data

□ HRコンサル

Human Resourcesについてのデータ分析
→ コンサル業務への素材提供の意味合いも・・・

最近では、金融に限らず機械学習色の強いテーマにも取り組んでいる

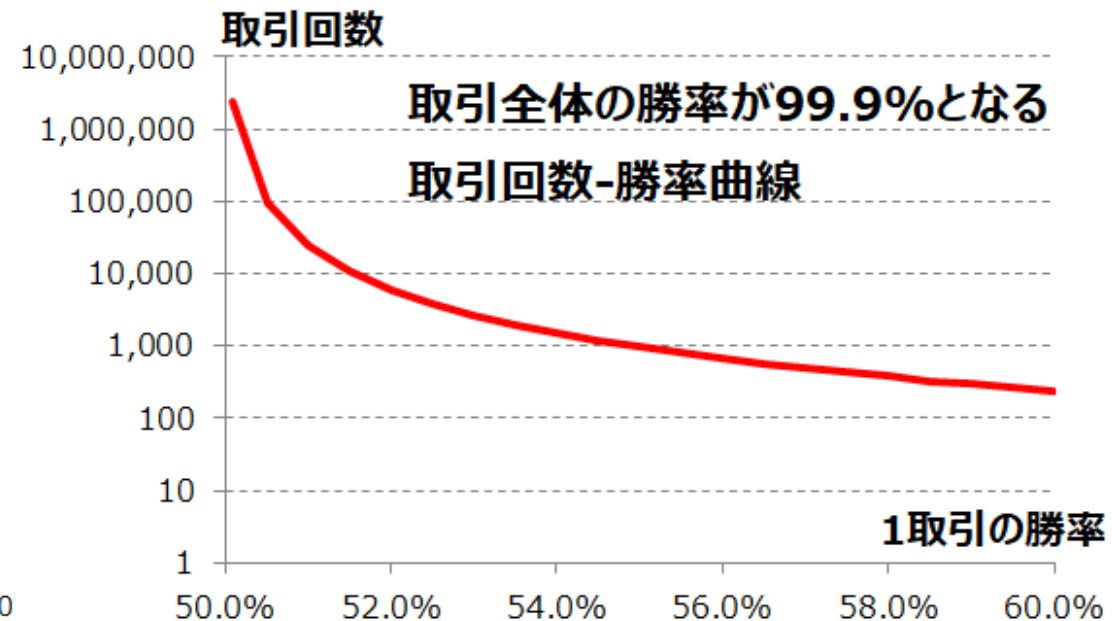
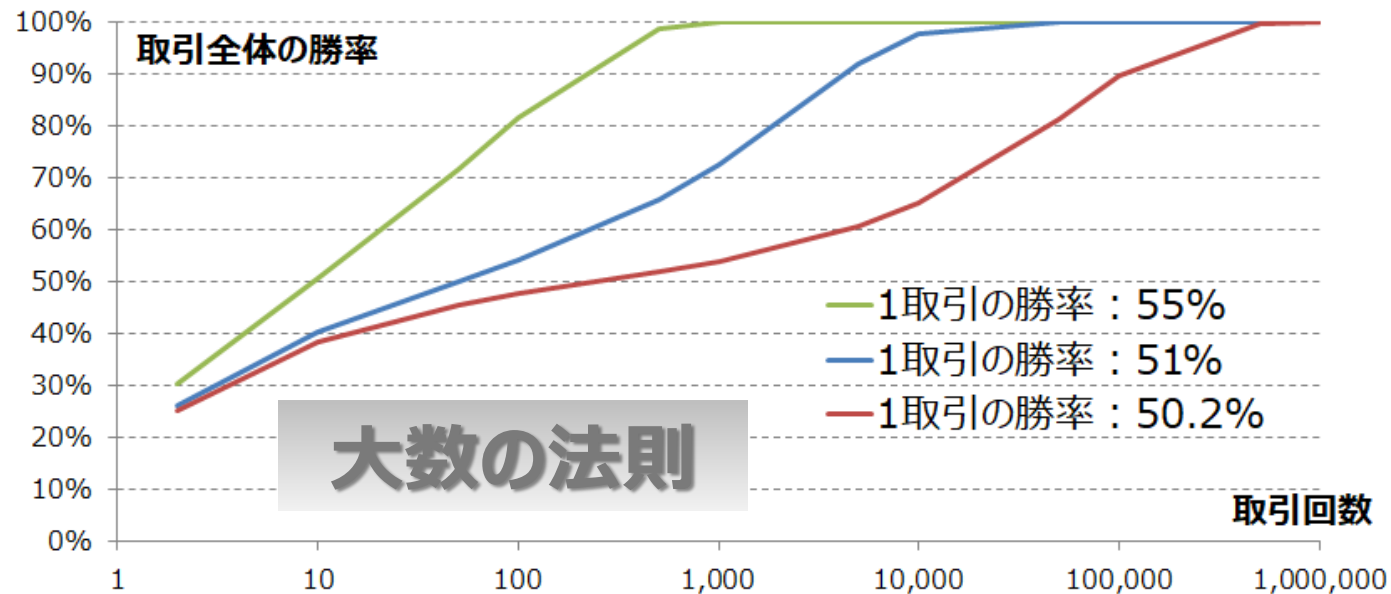
- Economic News Impact Analysis Using Causal-Chain Search from Textural Data, **AAAI-20 Workshop**（東大との共同研究）
- Output Diversified Initialization for Adversarial Attacks, **ICLR-2020 Workshop**（スタンフォード大との共同研究）
- 高頻度金融データとGANを用いたデータ拡張, JAFEE2020夏季大会（都立大との共同研究）
- Diversity can be Transferred: Output Diversification for White- and Black-box Attacks, **NeurIPS 2020**（スタンフォード大との共同研究）

機械学習・オルタナティブデータを使って感じる事:どう使えばいいのか？

□金融における機械学習・オルタナティブデータの使い方

表が出たら“1円貰う” 裏が出たら“1円払う”公正なコイン投げでは、いくら機械学習を用いても勝つことは出来ない

⇒では、どれだけ歪んだコインを何回投げれば勝率はどれくらいになるのか？



株式投資に置き換えると・・・

- 儲かる確率の高い株式を選ぶ (なるべく独立な方が望ましい)
- 売買する機会を増やす (これも独立で、取引コストが安いほうが望ましい)

のポイント

⇒これを実行しているのがHFT (高頻度取引)

大量の銘柄のTick Dataを機械学習等で分析し、リアルタイムな情報も加え、勝てそうな銘柄を探し出して高頻度に売買

▶人間には出来ないアルゴリズム対アルゴリズムの戦い

機械学習・オルタナティブデータを使って感じること:どう使えばよいのか？

□金融における機械学習・オルタナティブデータの使い方

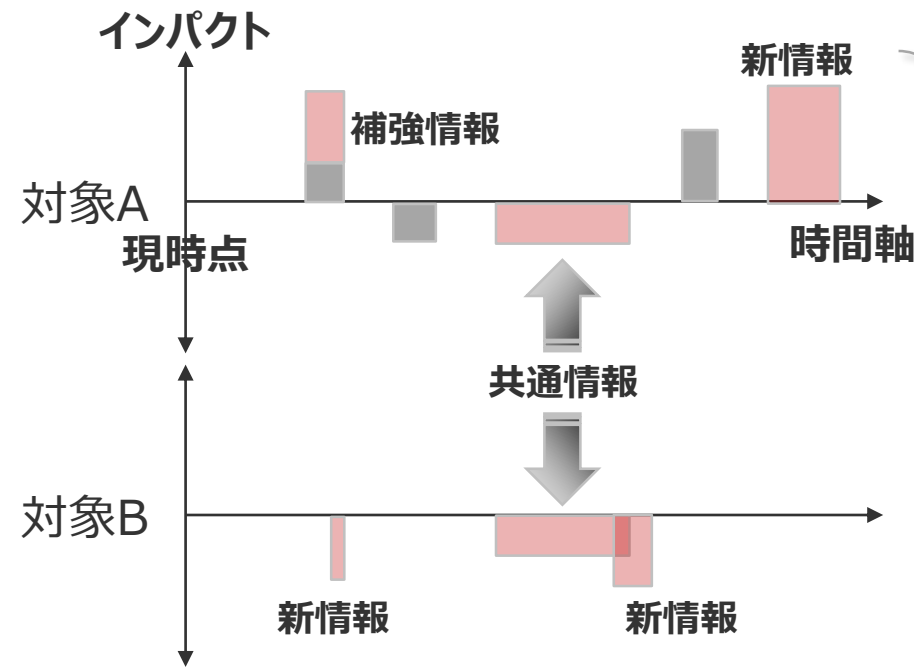
クオンツ的な機関投資家の視点で考えると、まずは価値のある**情報抽出**

- 儲かる確率の高い投資対象を選ぶ（なるべく独立な方が望ましい）

⇒ 既存データの**補強**や**新しい情報**で**確度向上**

- 売買する機会を増やす（これも独立で、取引コストが安いほうが望ましい）

⇒ **新しい評価軸**を提供し、投資の**対象範囲**や**頻度**を**拡張**



■ 既存の情報

■ オルタナデータ+機械学習による情報

□個別の情報に着目

⇒ 銘柄ピックアップ
イベントドリブン

□共通の情報に着目

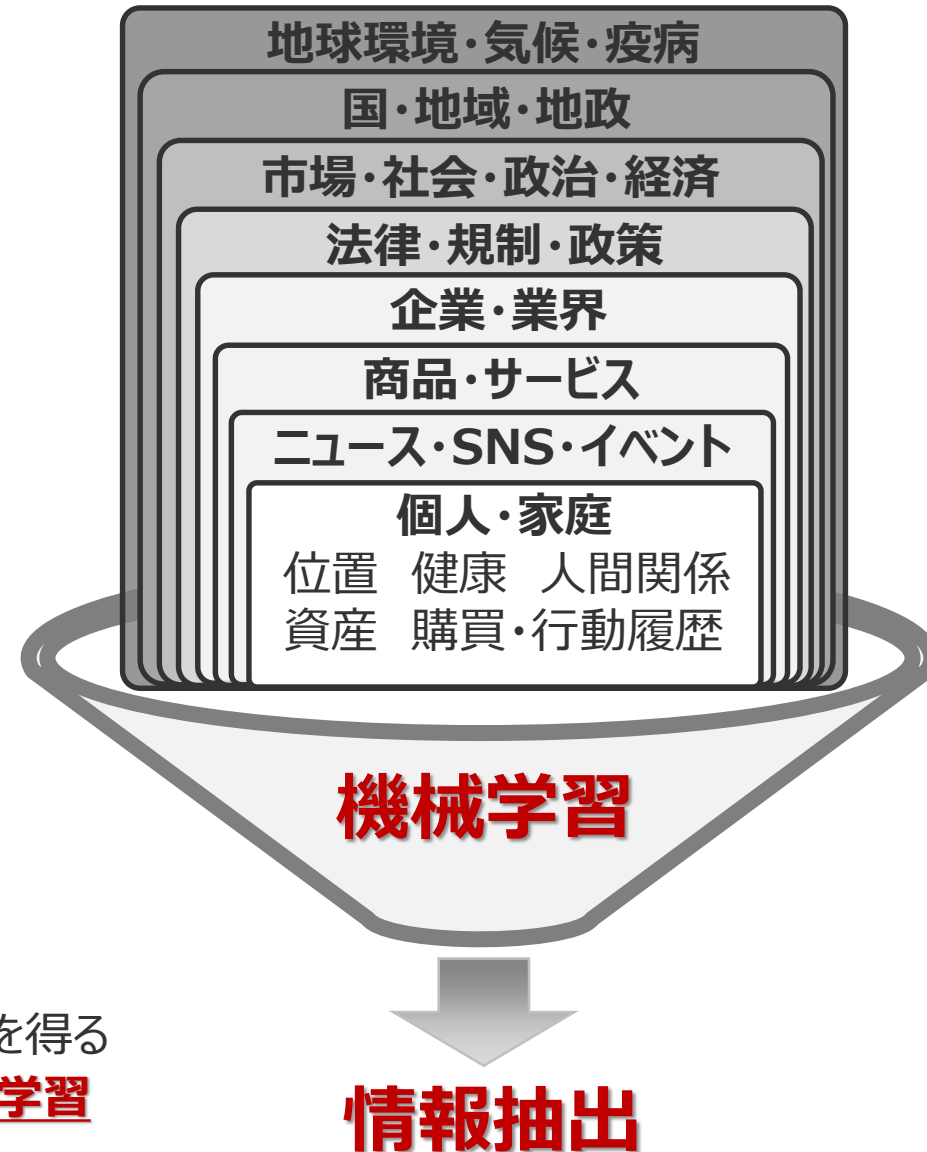
⇒ ファクター投資
インデックス投資
スマートベータ戦略

投資目的に応じて、適切な情報を得る

(オルタナティブ) データ×機械学習

の組み合わせを選択

▶ 環境変化に対応することも大事！



影響範囲 タイムホライズン インパクト

機械学習・オルタナティブデータを使って感じること:使い方の問題

□金融における機械学習・オルタナティブデータの使い方

株式の中長期運用のように「時間による不確実性が高い」領域では、まだ機械学習やオルタナティブデータが十分に活用されていないという認識

⇒ HFTのように環境変化の不確実性が低いマイクロな領域では活用されている

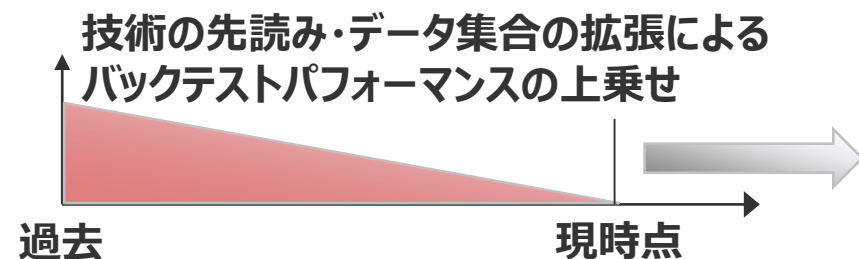
中長期の資産運用におけるバックテストは、いろんなイベントの影響を考慮するため10年～と長めに設定される
そうすると、無意識に機械学習やオルタナティブデータを使ってしまう・・・

- その時点では利用できなかったオルタナティブデータ・・・データ集合の**拡張** ▶ **情報が増えた！**
 - その時点では利用できなかった機械学習の手法・・・データを**効率的に分析** ▶ **情報を早く得た！精度もup！**
- を用いて、

●その時点で利用できるものだけで取引した結果であるベンチマーク

を上回ることが起きやすくなってしまう

これは、いわゆる**データの先読み**はしていないが、**技術の先読み・データ集合の拡張**はやってしまっている状態
これが**将来におけるパフォーマンス**を保証するに足るのか？ 静的な過去ではなく、今の投資家に対して優位か？



実務において重要な「本番での運用」での有用性をどう判断？

—ロバスト性（過学習の問題）—

—ロバスト性（過学習の問題）—

機械学習に限ったことではなく、クオンツ運用あるあるのひとつとして・・・

●バックテストではパフォーマンスがよかったのに、実際に運用を始めると残念な結果になってしまうということが良くあります。

原因のひとつとして、アルファや戦略を構築する際に、バックテストの結果を見ながら、様々なパラメータをチューニングしてしまうことが挙げられます。「バックミラーを見て運転する」と揶揄される行動ですが、同じコースを何回も運転すると何となく上手くなったと感じられるのと同じように、過去のひとつしかないヒストリカルデータで何度もバックテストすると、それに過剰にフィットしたモデルになってしまいます。（で、新しい環境・新しいデータでは事故ってしまう・・・）

そこで、

- GAN（Generative Adversarial Network：敵対的生成ネットワーク）等の手法で市場データと見分けのつかないデータを作り、多数のパスでロバスト性を確認
 - パラメータの変化に対するパフォーマンスの感応度を確認
 - ノイズから成るアルファと統計的に違うか？先読みデータをどれくらいノイズに混ぜ込んだものと同様か？等を確認
- などいくつかの対処法に取り組んでいます。

とは言え、アルファや戦略を構築した時と環境が変わったのか否かは、何を基準に判断すれば良いか難しい問題です。それに、もし環境変化が把握できたとしても、変化に対してすぐに対応するのも難しいかと思います。

—オルタナティブデータのコストパフォーマンス—

ーオルタナティブデータのコストパフォーマンスー

最終的にはファンドのアルファやリスク管理に使うので、オルタナティブデータの有無により、どれだけパフォーマンスが改善するかで判断することになります。

結構コストが高く、買うとなるとROIを問われて困るため、残高増加による手数料への貢献と天秤に掛ける感じです。また、

- データ提供の継続性が担保されているか？
- 途中でコストが増大するような虞はないか？
- 他の伝統的データでレプリケート出来ないか？
- 加工・作成方法が開示されているか？
- 簡単に扱えるか？

など、なるべく透明性があって、安く安定的に供給されることもポイントかと思います。

お勧めは・・・特定のデータではないのですが、データベンダーが多数（100以上）のオルタナティブデータを使えるようなサンドボックス環境を提供していますので、まずはそこで自分の目的に合うデータを探すところからスタートするのがコストパフォーマンス的には良いかと考えています。

ただ、特定ベンダーの独自技術やサービスにあまりに依存し過ぎると、値上げに抵抗できなかったり、その環境を使い続けなくてはいけなくなったり、他ベンダーへの乗り換えが困難になったり・・・とベンダーロックイン状態になることもあるので注意が必要です。

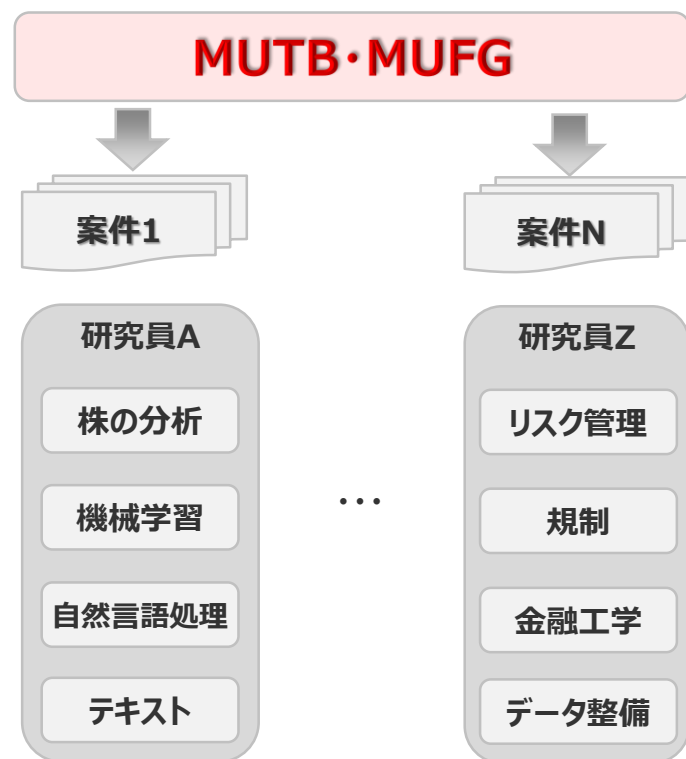
リサーチ体制

機械学習・オルタナティブデータを使って感じること:適した研究環境に変化を！

□これまで（縦割り型）

案件（研究テーマ）ごとに研究員を割り当てていたので、
新規技術・データに不慣れなひともいる・・・

⇒研究員のスキルに応じて成果や分析時間が変わる



煩雑な非構造データの処理・日進月歩の機械学習手法・・・

⇒オールマイティな研究員育成に時間が掛かる！

□これから（プラットフォーム型）

研究プロセスごとに適材適所の人材をアロケーションし、
組織力を活かした研究体制に移行

⇒コミュニケーションや共通化が鍵



先に組織としての専門性を高め、後々個人を高める

機械学習・オルタナティブデータを使って感じること:研究-開発-運用プロセスを効率化！

□研究プロセスの再定義（PDCAの高速化が必要！）

研究プロセスは研究-開発-運用（ResDevOps）サイクルの一部

ビジネスアナリティクス

金融実務 + 新商品・サービス提案力

ユーザーコミュニケーション
(要因分析)

ビジネスリサーチ
(金融/IT)

評価

課題定義

仮説生成

アカデミックリサーチ
(金融実証/理論)

アカデミックリサーチ
(ロバスト最適化)

モニタリング

研究プロセス

モデリング

データアナリティクス

金融理論 + 新モデル・分析提案力

仮説検証

データ準備

アカデミックリサーチ
(統計/因果推論)

データサイエンス

金融技術 + 新手法・テスト提案力

アカデミックリサーチ
(統計/機械学習系)

データマネジメント
データアーキテクチャ
データエンジニアリング
データガバナンス

□開発

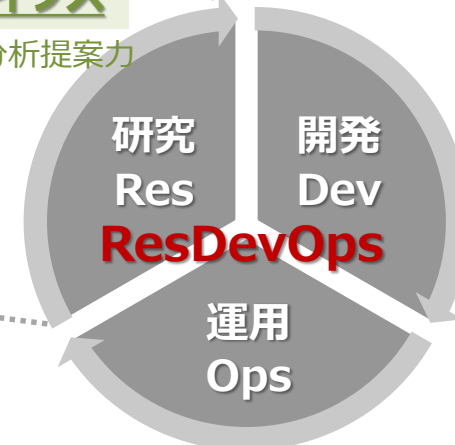
研究成果をスムーズにサービス提供に結び付ける

- ⇒ 機械学習等のモデルは短期間での入替え前提
- クラウドやオープンソースを利用して、疎結合でスケーリング可能なシステムを目指す！

□運用

業務が継続する間はサービスを提供し続ける

- ⇒ 自然言語処理で用いる辞書も適宜更新
- COVID-19のニュースをタイムリーにタグ付け
- 新しい会社名などの新語・未知語に対応
- 自動化しないと手作業はツライ...



—実務家として伝えたいこと—

ー実務家として伝えたいことー

（平穏バージョン）

高いオルタナティブデータを買ったから・・・機械学習のPoCに高い費用払ったから・・・と sunk cost の呪縛に捕らわれて、実務家のクオンツがダークサイドに落ちていくことのないように、新しい技術の活用方法について、アカデミアの皆さまから様々な意見をお伺いできればと考えておりますので、共同研究も含めて積極的に交流させていただければ幸いです。現場には研究の元ネタになりそうなものもありますので・・・

（不穏バージョン）

ファイナンスは社会科学の一部門（実学でもある？）なので、過去のデータを用いた精密な分析も大事だとは思いますが、自然科学や工学と違い、再現性が乏しい分野な気がします。なので、環境変化に対応して変わっていく人の行動は、過去を深く掘り下げても得られるものは少ないのかもしれませんが。機械学習もそうですが、データが無いから使えない、ではなく、人々のこれからの不安や課題をどうやって解決していくか？に、ファイナンス的・定量的なアプローチをどのように活用していくかが今後問われるのかな～と思います。

（居酒屋バージョン）

これからの日本は人口動態的にも世界初の未体験ゾーンに突入していきます。過去のデータ分析が将来の何かを保証する割合が少なくなるかもしれません。中長期的に見て、ファイナンス分野での（学問の世界だけではなく）日本としての課題は何なのか？みんなが解決して欲しいと思っていることは何なのか？を、研究者としての探求心を追い求めると同時に、学会としても課題リストを整備し、社会からの期待に応える必要も（何割かは）必要となるのかもしれません。実務家はお客様から叩かれ慣れてますので問題ないですが、アカデミアの皆様はストレスに感じるのかもしれませんが・・・