

---

日本銀行決済機構局・金融市場局合同コンファレンス「AIと金融サービス・金融市場」  
第2部：パネルディスカッション「AIと金融市場へのインパクト」討議資料

## ビッグデータ／機械学習のアクティブ運用への利用

2017年4月13日

ゴールドマン・サックス・アセット・マネジメント株式会社

金融商品取引業者 関東財務局長(金商)第325号

加入協会 日本証券業協会、一般社団法人投資信託協会、一般社団法人日本投資顧問業協会、一般社団法人 第二種金融商品取引業協会

---

**ビッグデータ／機械学習の活用により、人間が行っていた投資判断をコンピュータ上に再現することが可能になりつつある。**

- 従来の技術では、経営者自体の評価、経営の質的側面を評価することは困難だった
- しかし、経営者の発言や過去の経営判断を分析し、経営者の行動や性格を評価することは可能になっている

## 従来からのクオンツ運用からの変化

- 構造化データ(数値)以外の情報を利用
- 数値・統計処理に加えて、自然言語解析(Natural Language Processing)、機械学習などの技術を利用
- クラウドコンピューティング、データ分散処理の多様により、テクノロジーへの依存が強まる
- 求められる人材が変化

## リスク

- 人工知能(AI)が万能であるとの誤解
- ビッグデータ／AIを使うこと自体が目的化すること

# 株式投資で利用可能なビッグデータ

## 構造化データと非構造化データ

### 構造化データ

- 財務データ、株価、アナリスト業績予想のように処理がしやすい数値データ
- 新しい構造化データも出現
  - 例: 携帯電話の位置情報、クレジットカード、ウェブ、業種特化

### 非構造化データ

- 文章、画像、音声など、特定の構造定義を持たないデータ
- 非財務情報を含む文章データの解析に多くの実用例
  - 企業: アニュアルレポート、10-K、10-Q、プレスリリース、決算説明会
  - メディア: ニュース、アナリスト・レポート
  - インターネット: ソーシャルメディア、掲示板、金融系ウェブサイト
- 画像の利用では、衛星写真を独自のアルゴリズムで分析し投資情報を提供するフィンテックも複数登場
- 音声認識などもすでに実用化されていると見られている

## リレーショナル・データベース以外のデータベース管理システム／データの分散処理技術

- 1箇所の集中管理データベースでデータを蓄積させるのではなく、複数の場所でビッグデータを蓄積させる分散型処理技術の発達により、膨大なデータの並列処理が可能になり、データ処理時間が短くなるとともに、大量のテキストデータの処理と管理が行えるようになった。

## クラウド・コンピューティング

- オンプレミスの計算資源を必ずしも所有しなくてもよくなり、必要に応じて計算資源を随時調達出来るようになった。

## 機械学習アルゴリズム

- 教師あり機械学習
  - 過去のデータ(トレーニング・データ)を分析し、入力データとラベル付けされた出力データの間関係をモデル化
  - 例えば、手書き文字認識では多数の手書きの文字を観察するだけでなく、そこに含まれる文字について事前にラベル付けされている。機械学習アルゴリズムは入力データ(文字の画像、色、文字の濃さ)と出力データ(実際の文字)の間関係を学習する。
- 教師なし機械学習
  - 構造化のために大規模な入力データを分析し、分類する。
  - たとえば、TシャツのサイズをS, M, Lをどのように設定するとき、目標顧客層の身長・体重のデータをもとに、アルゴリズムは身長と体重の分布を分析し、類似する身長と体重の層を含むS, M, Lの3つのセグメントを作成する。

上記はビッグデータについて理解を深めるための一般的な情報提供を目的としたものです。

# ビッグデータ活用の課題と展望

データベース開発、品質管理、人材とモデルの構造



Asset  
Management

ビッグデータを活用しようとする、コンピュータ・ハードウェアやソフトウェアなどの技術以外にもクリアすべき課題は多い

## データベース開発

- 金融以外で利用することを想定したデータを使う際は、データの二次加工が必要
- 非構造化データを使うときはデータベース格納前にメタデータ(追加情報)を付与するなどの前処理が必要で、これ自体が機械学習モデルである場合が多い

## 品質管理

- 評価対象企業と、データが必ずしも対応していない
- データ品質を管理するための専用プログラムやデータアナリストを配置する必要

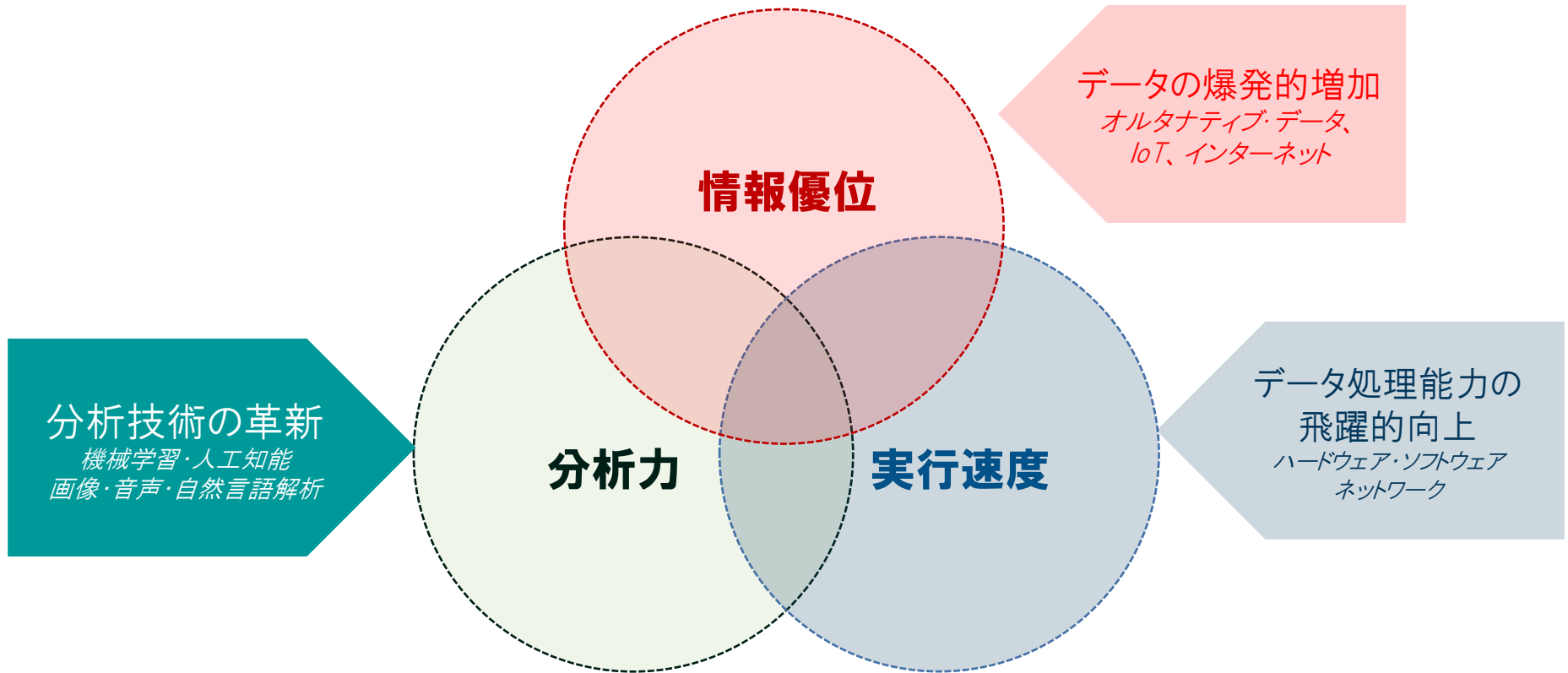
## 人材

- 自然言語解析、機械学習などの金融分野以外のデータ分析技術が必要、従来のクオンツアナリストだけでは対応できない
- データサイエンティストを新たに雇用する必要があるが、人材が不足している

## モデルの構造

- データが膨大なため、一度にまとめてデータ処理するバッチ型プロセスでは処理が間に合わない可能性も
- データが生み出される速度が速いため、処理の高速化、メモリーデータベースの必要性も

# ビッグデータ時代の投資戦略の優位性とは



上記は、例示をもって理解を深めていただくことを目的とした概念図です。  
上記はビッグデータについて理解を深めるための一般的な情報提供を目的としたものです。

- 本資料は、情報提供を目的としてゴールドマン・サックス・アセット・マネジメント株式会社(以下「弊社」といいます。)が作成した資料であり、特定の金融商品の推奨(有価証券の取得の勧誘)を目的とするものではありません。
- 本資料は、弊社が信頼できると判断した情報等に基づいて作成されていますが、弊社がその正確性・完全性を保証するものではありません。
- 過去のデータは将来の動向を示唆あるいは保証するものではありません。
- 過去の運用実績は将来の運用成果を示唆あるいは保証するものではありません。投資価値および投資によってもたらされる収益は上方にも下方にも変動します。この結果、投資元本を割り込むことがあります。
- 本資料に記載された見解は情報提供を目的とするものであり、いかなる投資助言を提供するものではなく、また個別銘柄の購入・売却・保有等を推奨するものでもありません。記載された見解は資料作成時点のものであり、将来予告なしに変更する場合があります。
- 個別企業あるいは個別銘柄についての言及は、当該個別銘柄の売却、購入または継続保有の推奨を目的とするものではありません。本資料において言及された証券について、将来の投資判断が必ずしも利益をもたらすとは限らず、また言及された証券のパフォーマンスと同様の投資成果を示唆あるいは保証するものでもありません。
- 本資料に記載された経済、市場等に関する予測は、資料作成時点での様々な仮定や判断を反映するものであり、今後予告なく変わる可能性があります。これらの予測値は特定の顧客の特定の投資目的、投資制限、税制、財務状況等を考慮したものではありません。実際には予測と異なる結果になる可能性があり、本資料中に反映されていない場合もあります。これらの予測は、将来の運用成果に影響を与える高い不確実性を伴うものです。したがって、これらの予測は、将来実現する可能性のある結果の一例を示すに過ぎません。これらの予測は一定の前提に基づく推定であり、今後、経済、市場の状況が変化するのに伴い、大きく変わることが考えられます。ゴールドマン・サックスはこれら予測値の変更や更新について公表の義務を有しません。

## 守秘義務

- 本資料の一部または全部を、弊社の書面による事前承諾なく(Ⅰ)複写、写真複写、あるいはその他のいかなる手段において複製すること、あるいは(Ⅱ)再配布することを禁じます。
- © 2017 Goldman Sachs. All rights reserved. <89076-OTU-504861>