

プレゼン資料

2017年1月26日

アークス研究所
田中 久司

目次

自己紹介

1. 将来予測のためのミニAI開発

2. AI（機械学習）のツールとプログラミング

3. 米国 AI ベンチャーの調査

3.1 業界マップ

3.2 ビジネスモデル

4. ツールや手法の研鑽（Kaggle等のコンペへの参加）

自己紹介

一昨年リタイア、その前は、
総合商社で、投資案件の審査と全社リスクマネジメントを担当
その後、上場会社の監査役等を経て、現在は、リスクマネジメントとAIを研究

投資案件の審査とリスクマネジメントで痛感したのは、将来を読むことの難しさ

そこで一時期シナリオプランニングを研究し、
鳥の眼、虫の眼、で俯瞰することの重要性を認識

昨年より、AIブームに触発されて、AIが将来を読むことに使えないかをテーマに研究を続けている。

1. 将来予測のためのミニAI開発（やりたいこと）

目標

ミニAIの開発にトライするが、作りたいものは、Global Analytics（世界のマクロ経済動向を予測する計算知識エンジン）で、その開発の準備として、

1. AI（機械学習）のツールとプログラミング

2. 米国AIベンチャーの調査

を行った。

2. AI（機械学習）のツールとプログラミング

実際に自らがAI開発のトライアルを行い、その知見を深める。

■ やって見たこと

データ	やること	ツール名	言語、ライブラリ、分析ロジックなど
MNIST 手書き 数字	0～9の数字の判別 (画像認識)	Chainer	- Python(scikit-learn)、 - Chainer
		クラウド系	- Ruby（画像データの数値データへの変換） - 多項ロジスティック回帰分析
		AWS (Amazon)	多項ロジスティック回帰分析
		Azure (Microsoft)	- Python（数値データのjpegへの変換） - Watsonの画像認識(Visual Recognition/ Custom Classifier)
Kaggle 世界の マクロ データ	日本の GDP伸び 率の予測	Python	Python (numpy, pandas, matplotlib, seaborn, statsmodels, scikit-learn(linear_model.LinearRegression(), AdaBoost), XGBoost)

次に、トライした内容と結果の要点を説明します。

前ページから続く

単回帰および重回帰分析の実施
(Pythonのstatsmodelsライブラリ)

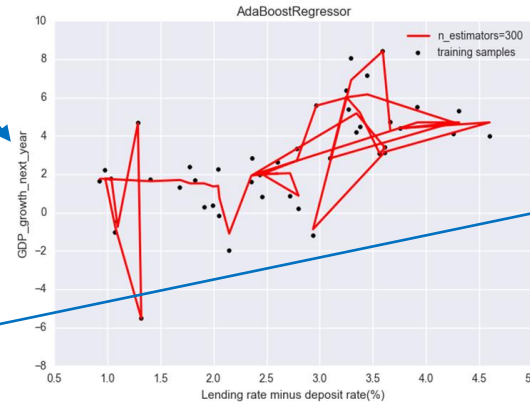
AdaBoostによる回帰分析
(PythonのScikit-learnの
DecisionTree Regression with
AdaBoost)

XGBoostによる回帰分析
(XGBoostのPython API)

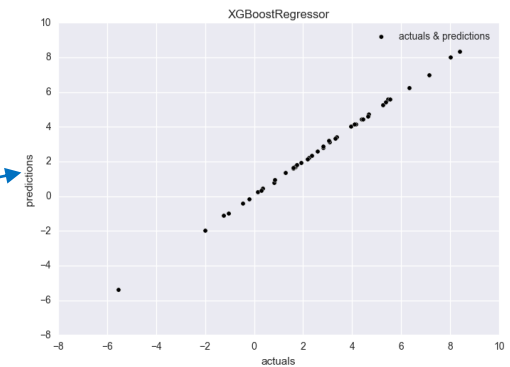
多変数の時系列データにおいて、
ある変数の将来値を予測するプロ
グラムの一応の雛型ができた。
プログラムにしても結果の解釈に
しても、アルゴリズムにより内容
が異なり、さらに習熟すべき課題
はあるが、最先端のXgBoostも動
かすことができ、結果も従来の回
帰分析よりも良好であった。

分析	正相関の説明変数	負相関の説明変数	R ²
単回帰	Lending rate minus deposit rate(%)	-	0.654
重回帰 (1)	Lending rate minus deposit rate(%)	Energy use(kg of oil equ./capita)	0.694
重回帰 (2)	上記+Land under cereal production (ha)	上記+ Imports of goods&services(yen)	0.434
重回帰 (3)	上記+Population, ages 0-14 (% of total)	上記+Exports of goods&services(yen)	0.437

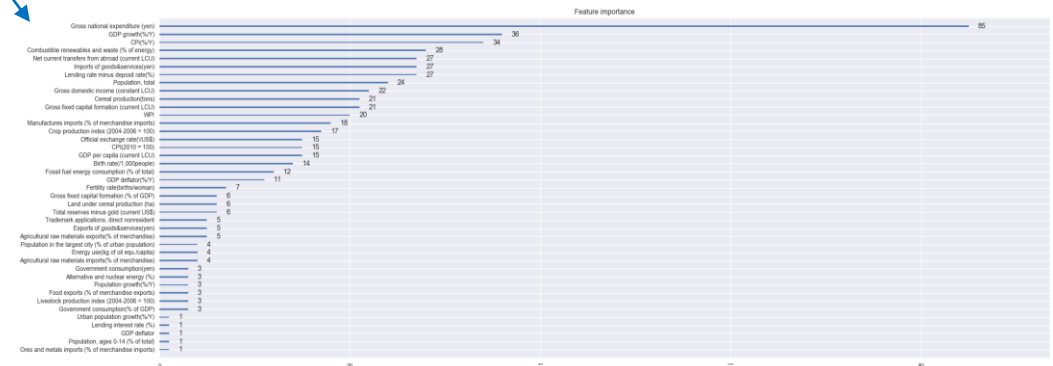
回帰分析の結果



Lending rate minus deposit rate(%)を説明変数とするAdaBoostによる単回帰分析



XGBoostによる回帰分析
(予測値と実際の値)



XGBoostによる各データの重要度

■ AI開発の勘所

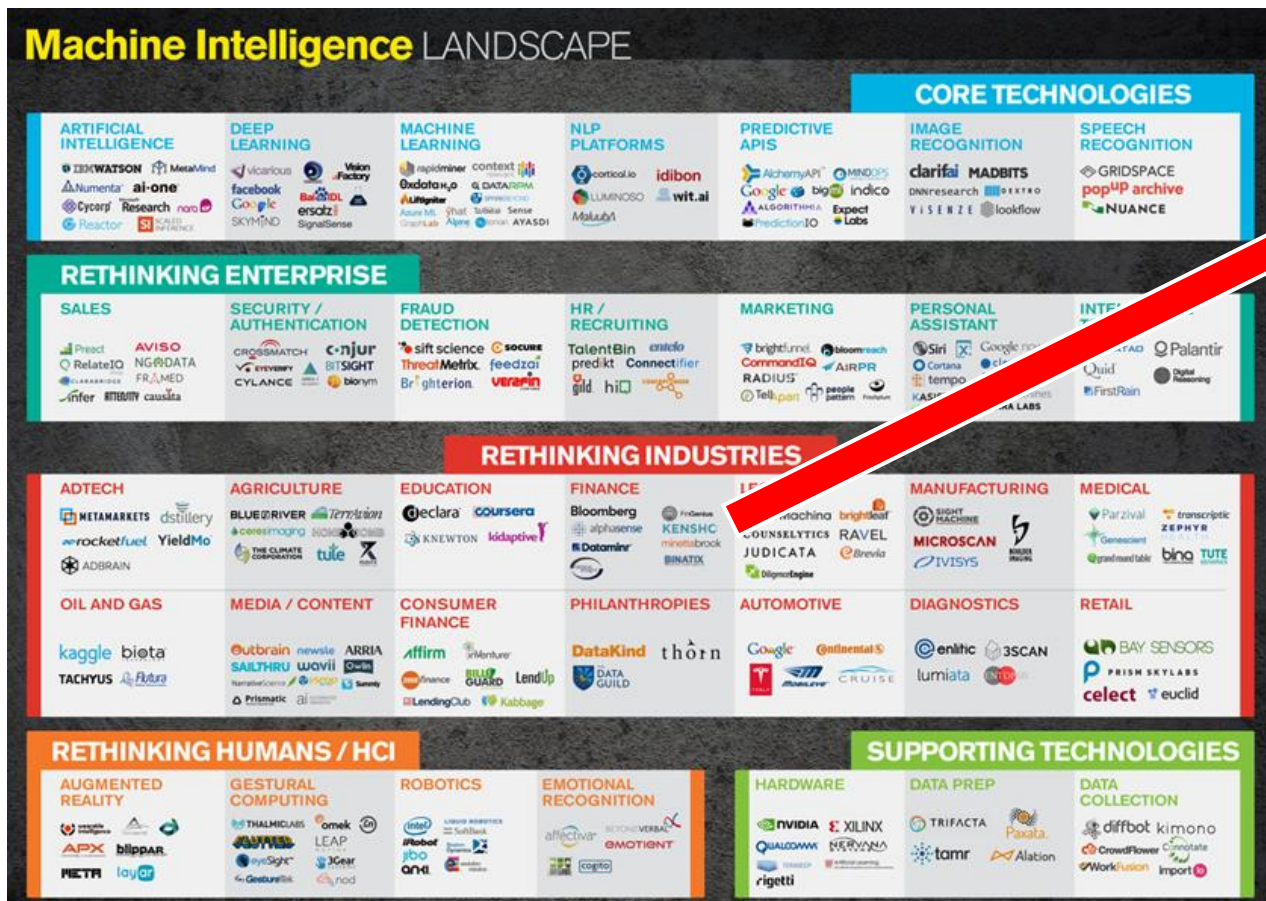
視点	勘所
開発作業	データの取得、ツールの環境構築、データの整形に手間がかかる。勿論アルゴリズムの選択がキーではあるが、データをツールに読み込ませる型や形式が限定されており、それを理解する必要がある。時系列データにおいて、ある変数の将来値を予測するプログラムの一応の雛型ができた。 オープンソースのみを活用することによって、AI的なものが開発できることが判った。プログラミングにしても結果の解釈にしても、さらに習熟すべき課題はあるが、最先端のXgBoostも動かすことができ、結果も従来の回帰分析よりも良好であった。
ツールの選択	オンプレミス系のPython（スクラッチ開発とChainer）は機械学習の本質的学習（統計学など）が必要である。クラウド系は、多彩な機能のそれぞれの使い方を学ぶのに手間がかかるが、慣れれば複雑なプログラミングをせずに結果が得られる点は便利である。ユーザー企業の置かれた立場・目的によりツールの使い分けは決めるべきであろう。例えば、優秀なプログラマーがおり、本格的にAIに取り組むのであれば、Pythonによるスクラッチ開発、開発目的が特定され、それに適したライブラリがあるのであればRを選択するなどである。

視点	勘所
クラウド系への感想	クラウド系は、ITの巨大企業が巨額の先行投資を行って開発したクラウド・プラットフォームで、日々進化し、数多くの機能が盛り込まれているので、その全貌を理解するのは相当の努力が必要であり、その優劣は簡単には判らない。従って、他の業務で利用しているクラウドがあれば、それを使うのが無難であろう。 特に使い慣れたものがない場合は、AWSはクラウドサービスとしては先行しており薦められるが、機械学習機能が簡単な分類や回帰分析だけであるため、本格的な分析を行う場合は他のサービスやPython等によるスクラッチでの開発を検討する必要がある。 Azureは一通り機械学習を学ぶには体系化されており、キャンバスでのGUIは習得しやすく、フリーワークスペースが無期限で使えるのも有り難い。ただし資料は英語版に限定されている。 BlueMixは多様な機能が集約されており様々なニーズに対応できるが、夫々の使い方に習熟するにはそれなりの努力が必要である。資料は日本語版もあるが、詳細な説明となると英語版しかない場合がある。チャット機能によるサポートが得られるが、無料アカウントが1か月に限定されているのはつらい。

3. 米国 AI ベンチャーの調査

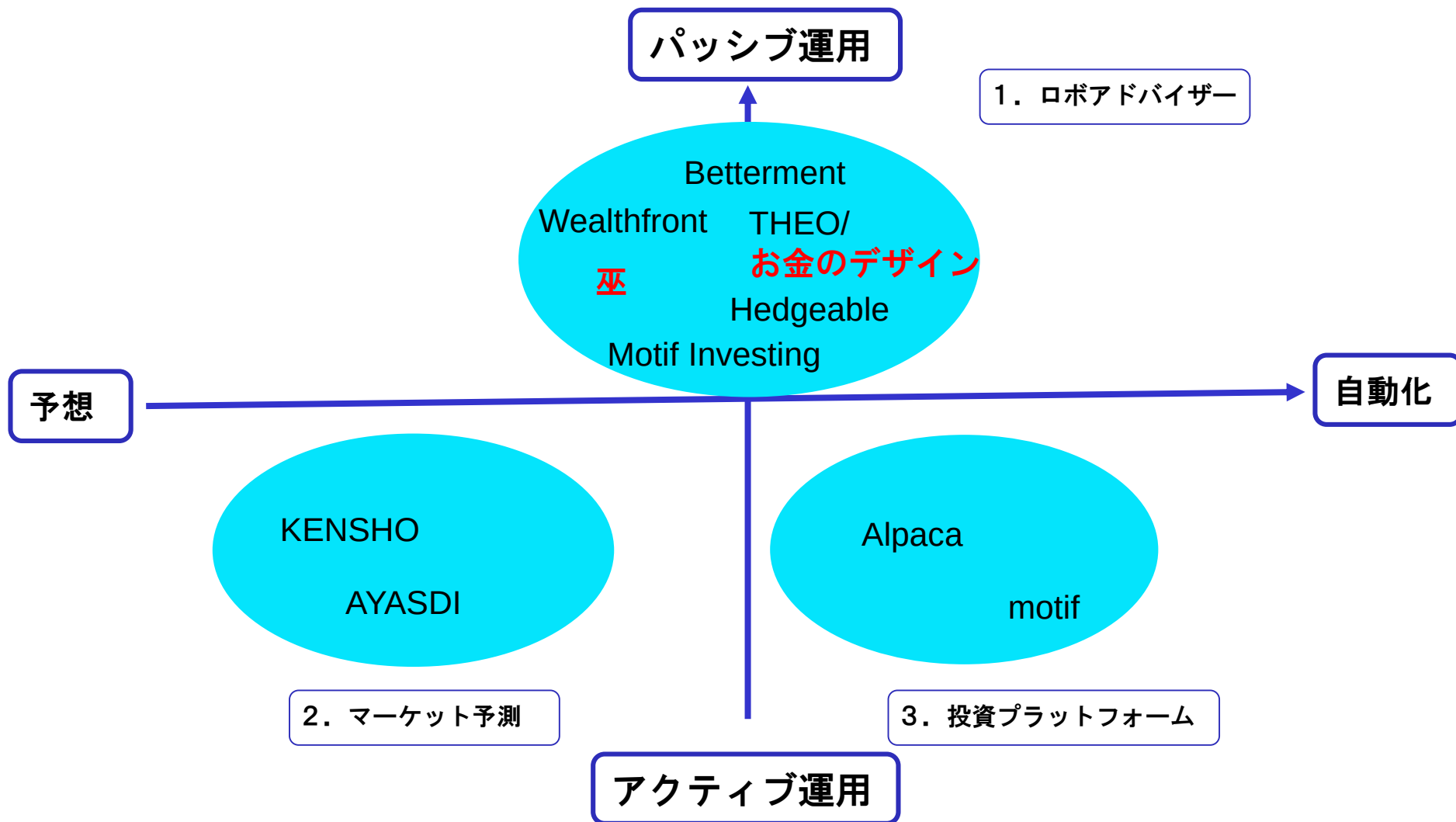
似た業態のベンチャーがあるかどうか調査する。

Machine Intelligence LANDSCAPE (ブルームバーグ社アナリストによる人工知能ベンチャーの調査結果)



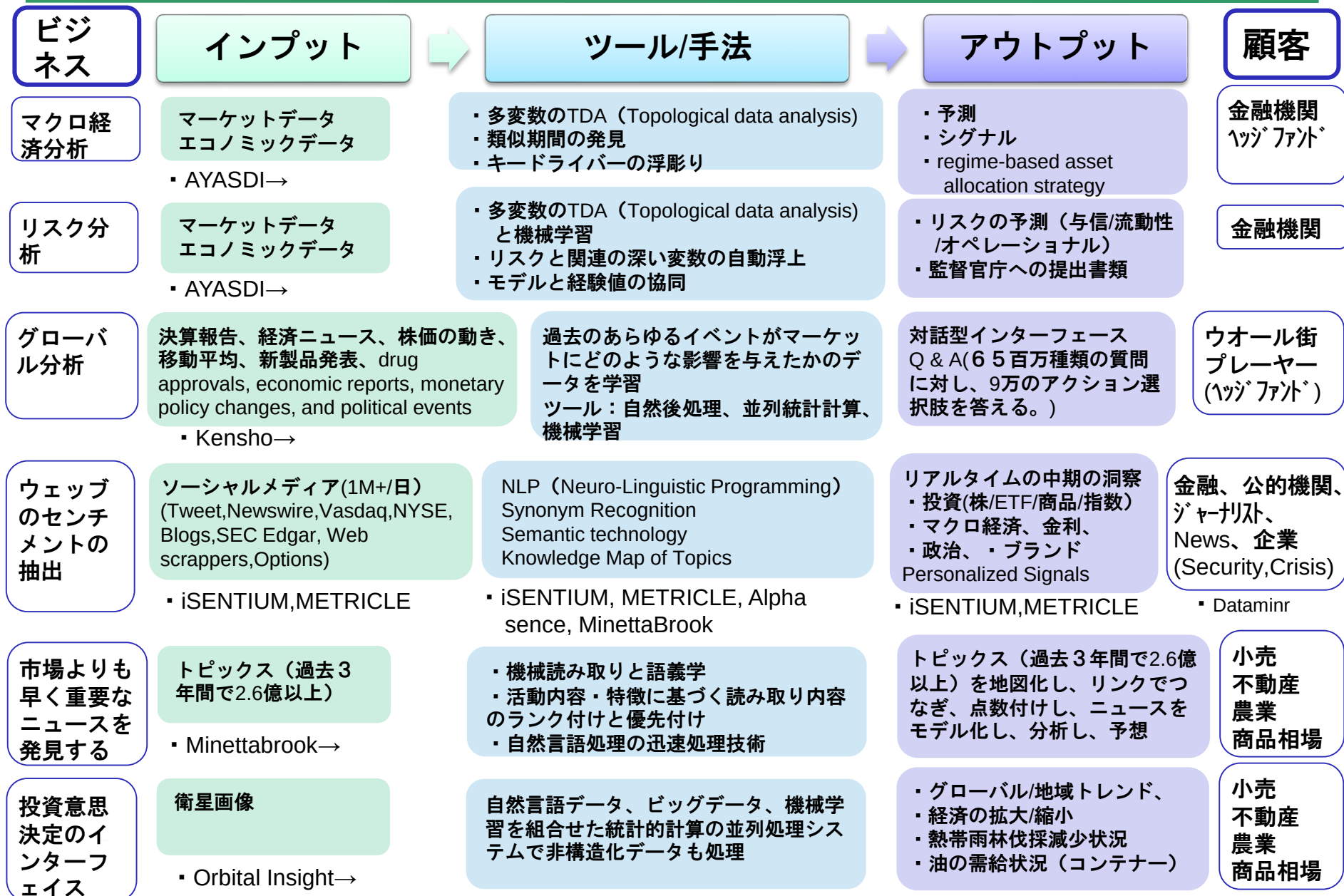
調査業種	調査社数
財務(投資)	9社
消費者金融	7社
小売	3社
製造	3社
医療	7社
診断・診察	3社
メディア	3社
ITテクノロジー	3社
オイル・ガス	5社
農業	2社
合計	45社

3.1 米の資産運用ベンチャーの業界マップ



- 【出典】 ①「最新Fin Techの基本と仕組みがよ〜くわかる本」長橋賢吾 著 (株)秀和システム 2016年12月10日発行
②「ロボアドバイザーの資産運用革命」 編著者 お金のデザイン (社)金融財政事情研究会 平成28年11月10日
③「Fintechとは何か」 編著者 富士通総研 隈本正寛・松原義明 (社)金融財政事情研究会 平成28年5月6日
④「These are the 5 Hottest Companies in Fintech」Fortune/Finance Jun 27,2016

3.2 ビジネスモデル



4. ツールや手法の研鑽（Kaggle等のコンペへの参加）

コンペ名	内容	To Go	懸賞金	チーム数
Data Science Bowl 2017	Can you improve lung cancer?	3 months	\$1,000,000	6,190
The Nature Conservancy Fisheries Monitoring	Can you detect and classify species of fish?	3 months	\$150,000	1,206
Dstl Satellite Imagery Feature Detection	Can you train an eye in the sky?	2 months	\$100,000	138
Two Sigma Financial Modeling Challenge	Can you uncover predictive value in an uncertain world?	A month	\$100,000	1,409

これにチャレンジ中