IEICE 情報理論研究会
若手研究者のための講演会

ビジネスデータを対象としたデータアナリティクスの現状と今後の展望

Current Development and Future Perspectives of Analytics for Business Data

早稲田大学 理工学術院 経営システム工学科

後藤 正幸

自己紹介

氏名：後藤正幸

所属：早稲田大学 創造理工学部 経営システム工学科

専門：情報数理応用、データサイエンス、経営情報分析
パターン認識と機械学習、情報理論、学習理論などの枠組みを、ビジネスデータ分析、マーケティング分析で活用するための方法論やモデルの研究に従事

学会：IEEE, INFORMS,電子情報通信学会, 情報処理学会, 人工知能学会,日本OR学会, 経営情報学会, etc.

その他：

早稲田大学 データサイエンス研究所所長

経営科学系研究部会連合協議会 データ解析コンペティションの運営メンバー

経 歴

- 1994年～2000年：早稲田大学 平澤茂一 研究室

  平澤茂一先生の研究室ご出身の大学人（一部）

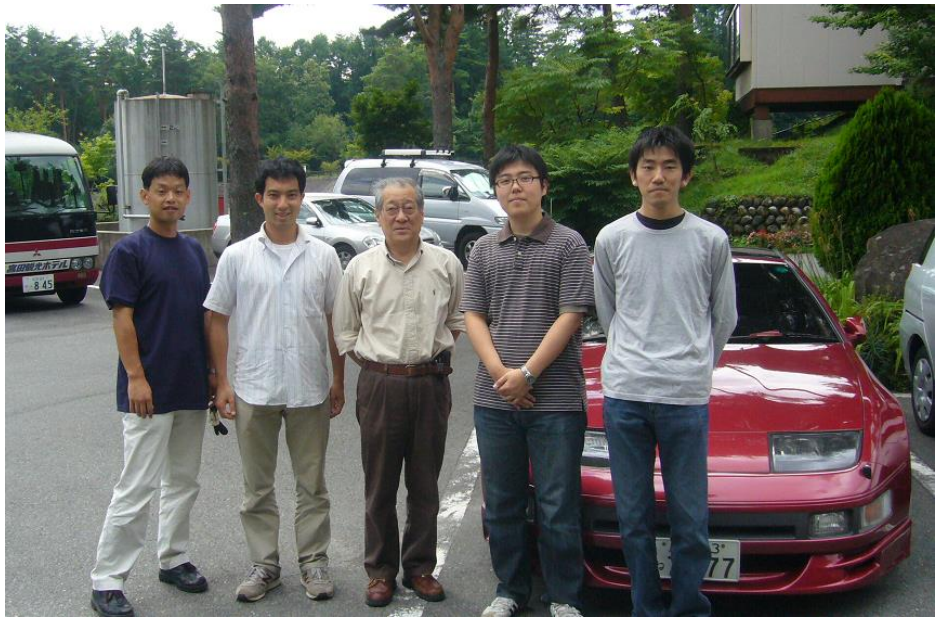
  - 稲積宏誠（青山学院大学）
  - 西島利尚（法政大学）
  - 松嶋敏泰（早稲田大学）
  - 鈴木 譲（大阪大学）
  - 松嶋智子（職業能力開発総合大学校）
  - 新家稔央（東京都市大学）
  - 鴻巣敏之（大阪電気通信大学）
  - 鈴木 誠（湘南工科大学）
  - 酒井哲也（早稲田大学）
  - 中澤 真（会津大学短期大学部）
  - 小林 学（早稲田大学）
  - 梅澤 克之（湘南工科大学）
  - 八木秀樹（電気通信大学）
  - 石田 崇（高崎経済大学）
  - 細谷 剛（早稲田大学）
  - 雲居玄道（早稲田大学）

- 2000年～2002年：東京大学 松島克守 研究室

- 2002年～2008年：武蔵工業大学 環境情報学部

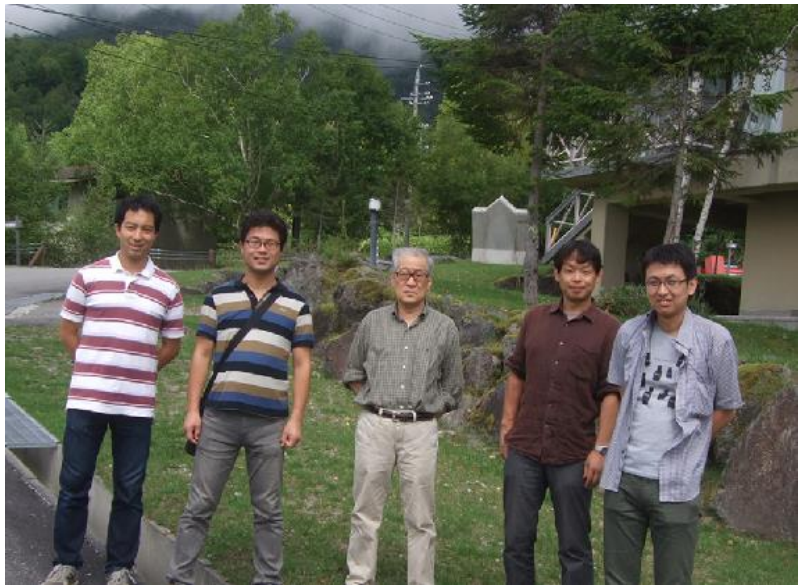
- 2008年～：早稲田大学 創造理工学部
  経営システム工学科

写真など（恩師・平澤茂一先生、同門の皆様と）



2008年8月 平澤研夏合宿

2010年10月 平澤先生の誕生日お祝い



2013年10月 平澤先生の別荘で

博士課程の頃の研究

- ベイズ統計の漸近的性質を研究していました

キーペーパー

B. S. Clarke and A. R. Barron, “Information-theoretic Asymptotics of Bayes Methods,” IEEE Trans. Inform. Theory, vol. 36, pp. 453–471, May 1990.

ベイズ符号

$$L_{Bayes}(x^n) = -\log \int_\theta P(x^n|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$E\left[L_{Bayes}(x^n)\right] = -\log P(x^n|\boldsymbol{\theta}^*) + \frac{k}{2}\log\frac{n}{2\pi e} + \log\frac{\sqrt{\det I(\boldsymbol{\theta}^*)}}{f(\boldsymbol{\theta}^*)}$$

MDL符号

$$L_{MDL}(x^n) = \min_\theta \{-\log P(x^n|\boldsymbol{\theta}) - \log f(\boldsymbol{\theta})\}$$

ある気付き

$$L_{Bayes}(x^n) = -\log \int_\theta P(x^n|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$L_{MDL}(x^n) = \min_\theta \{-\log P(x^n|\boldsymbol{\theta}) - \log f(\boldsymbol{\theta})\}$$

ベイズルール

$$f(\boldsymbol{\theta}|x^n) = \frac{P(x^n|\boldsymbol{\theta}) f(\boldsymbol{\theta})}{P(x^n)}$$

$\int_\theta P(x^n|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}$

$$\log f(\boldsymbol{\theta}|x^n) = \log \{P(x^n|\boldsymbol{\theta}) f(\boldsymbol{\theta})\} - \log P(x^n)$$

$$L_{Bayes}(x^n) - L_{MDL}(x^n) = \max_\theta \log f(\boldsymbol{\theta}|x^n)$$

ある気付き

$$L_{Bayes}(x^n) - L_{MDL}(x^n) = \max_{\theta} \log f(\boldsymbol{\theta}|x^n)$$

$$\boldsymbol{\xi} = \sqrt{n}\left(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\right)$$

$$f_\xi(\boldsymbol{\xi}|x^n) = \frac{1}{\sqrt{n}^k} f(\boldsymbol{\theta}|x^n)$$

$$f_\xi(\boldsymbol{\xi}|x^n) \to \frac{\sqrt{\det I(\tilde{\boldsymbol{\theta}})}}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{\xi}\|^2_{I(\tilde{\theta})}\right\}$$

この事後確率密度の漸近正規性で、ベイズ符号とMDL符号の漸近的な差は解析できる

- Masayuki Goto, Toshiyasu Matsushima, and Shigeichi Hirasawa : "An Analysis of the Difference of Code lengths Between Two-Step Codes Based on MDL Principle and Bayes Codes" , *IEEE Trans. Information Theory*, Vol.47, No.3, pp.927-944, 2001年3月
- Masayuki Goto, Toshiyasu Matsushima, and Shigeichi Hirasawa : "Almost Sure and Mean Convergence of Extended Stochastic Complexity", *IEICE Trans. on Fundamentals*, Vol.E82-A, No.10, pp.2129-2137, 1999年10月
- Masayuki Goto, Toshiyasu Matsushima, and Shigeichi Hirasawa : "A Generalization of B.S.Clarke and A.R.Barron's Asymptotics of Bayes Codes for FSMX Sources", *IEICE Trans. on Fundamentals*, Vol.E81-A, No.10, pp.2123-2132, 1998年10月

現在の研究活動

早稲田大学データサイエンス研究所

研究所名：データサイエンス研究所
設置期間：2015年10月1日～
研究所員：後藤正幸、守口 剛、大野高裕、
永田 靖、上田雅夫、鈴木広人、他

研究内容：
経営判断、マーケティングに関わるデータの利用促進に関する研究

協力企業：
マクロミル、ZOZO、良品計画、ヴァリューズ、
小田急電鉄、小田急エージェンシー、ラボラティック

経営科学系研究部会連合協議会
ー データ解析コンペティション

1994年（平成6年）より開催されている、「共通の実データをもとに、参加者が分析結果を競う」ことを目的とした取り組み

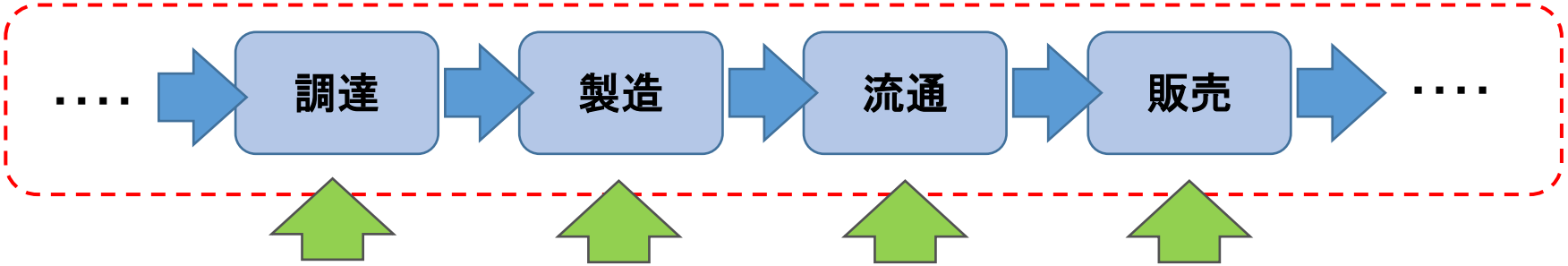
参加チーム数:　約110チーム
参加者総数：　延べ 700人

協議会を構成する学術団体

- 日本オペレーションズ・リサーチ学会，ビッグデータとマーケティング分析研究部会
- 日本マーケティング・サイエンス学会 ID付POSデータ活用研究部会
- 日本マーケティング・サイエンス学会 消費者・市場反応の科学的研究部会
- 日本マーケティング・サイエンス学会 消費者行動の学際的研究部会
- 日本マーケティング・サイエンス学会 市場予測のための消費者行動分析研究部会
- 日本計算機統計学会 データ解析スタディーグループ
- 日本データベース学会 ビジネスインテリジェンス研究グループ
- ACM SIGMOD 日本支部
- 日本経営工学会 経営情報部門

ビジネスアナリティクスを取り巻く現状

情報化時代の経営環境



IT　IT　IT　IT

EDI (Electronic Data Interchange) SCM (Supply Chain Management) ICチップ	MRP (Material Requirement Planning) SCM (Supply Chain Management)	EDI (Electronic Data Interchange) SCM (Supply Chain Management) 倉庫システム ICチップ	POS (Point of Sales system) ICカード

- あらゆる段階にITが入り込み、実質的に ITなしでは業務が出来ない現状
- ITを利用すると履歴が残る
  → 膨大な作業履歴情報の山

マーケティングリサーチと情報システム

ユーザの行動履歴は、全て情報システムを介して取得可能な時代
・・・　ICカードによる購買など

顧客の囲い込み

・電子マネー
・ポイント還元
・クレジット
・サービス授受
・商品推薦

マーケティングデータの取得

ICカードによるポイントシステム

- カードでモノが買える
- ポイントが付く

便利！！！

しかし……

小田急電鉄 クレジット機能付き OPカード

実は消費者のベネフィット
以上に、企業側にとって優れた
マーケティングツールとなっている

なぜか？

小田急電鉄 OPカード

ECサイトやポイントシステムによって蓄積される閲覧・購買履歴データ

- 消費者はインターネット上で自由に製品やサービスを閲覧し、購買行動を起こす。
- 消費者は購入製品の満足度を評価し、他のユーザ評価を参考にして、購買行動を起こす

マーケットリサーチの情報源

- ユーザ数：　数千～数百万
- アイテム数：　数十万～数百万

データマイニング、テキストマイニング
レコメンデーション

Googleトレンド

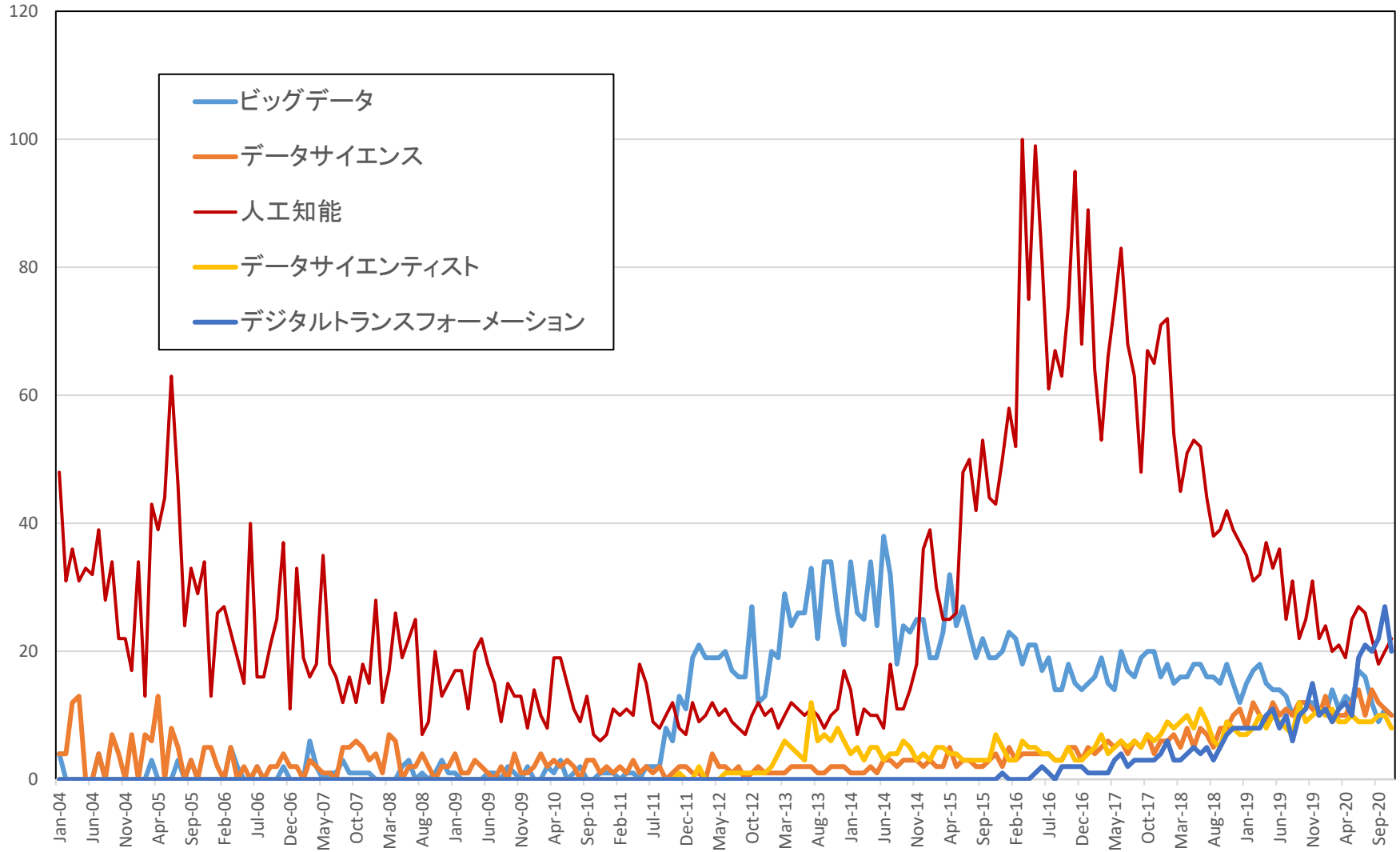
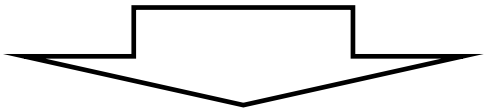


図. Googleトレンドによるキーワード推移

パターン認識と機械学習の現状

- ディープラーニングやWord2Vecの登場、囲碁や将棋のAIソフトの躍進などにより、人工知能ブームが再到来
- 大規模で多様なデータ（ビッグデータ）を取得・利用可能な情報環境が整備され、これまでになくデータ活用への期待が高まる。
- IoT（Internet of Things）、インダストリー4.0、ソサイエティ5.0など、新たな概念が脚光を浴びる。

一方で、、、



ビジネス領域におけるデータ活用は発展途上

ビジネスアナリティクス環境の変化

- ディープラーニングやWord2Vecの登場
- 囲碁や将棋のAIソフトの躍進などにより、人工知能ブームが再到来
- ビッグデータ処理の基盤技術(Hadoop, etc)
- 計算統計学の発達(潜在クラスモデル、ベイズモデル、etc)
- データマイニング、テキストマイニングのソフトウェアの充実

市販のPCでも相当量のデータの分析が可能となっている。

しかし、、、

- **分析の切り口が多様になり、結果の解釈が容易ではない。**
- **データはとにかく膨大であるが、分析しても何も出て来ない。**
- **様々なデータがあり過ぎて、何から手をつけて良いか分からない。**
- **一通り考えられる分析は終えたが、パッとしない。**
- **他のデータと組み合わせた分析が重要であると言われるが、具体的に何と統合して分析して良いのか分からない。**
- **データベース(IT)の扱いと統計分析の知識に長けた人材がいない。**

なぜ、ビジネスデータの分析は難しいのか？

Deep Learningが得意とする問題

- 高次元の特徴ベクトルであるが、人間が正解を与えられる問題
- 正解が変化しない問題
- カテゴリ同士に重なりが少ない問題

⇨ 従来のコンピュータには難しかったが、本来、人間は得意とする問題

ビジネスアナリティクス

- 人間によって正解が異なる問題
- 正解が状況によって変化する問題
- カテゴリ同士に重なりが多い問題

⇨ 人間にとっても 易しくはない問題

ビジネスアナリティクスにおける
パターン認識と機械学習

ビジネスアナリティクスで利用される機械学習

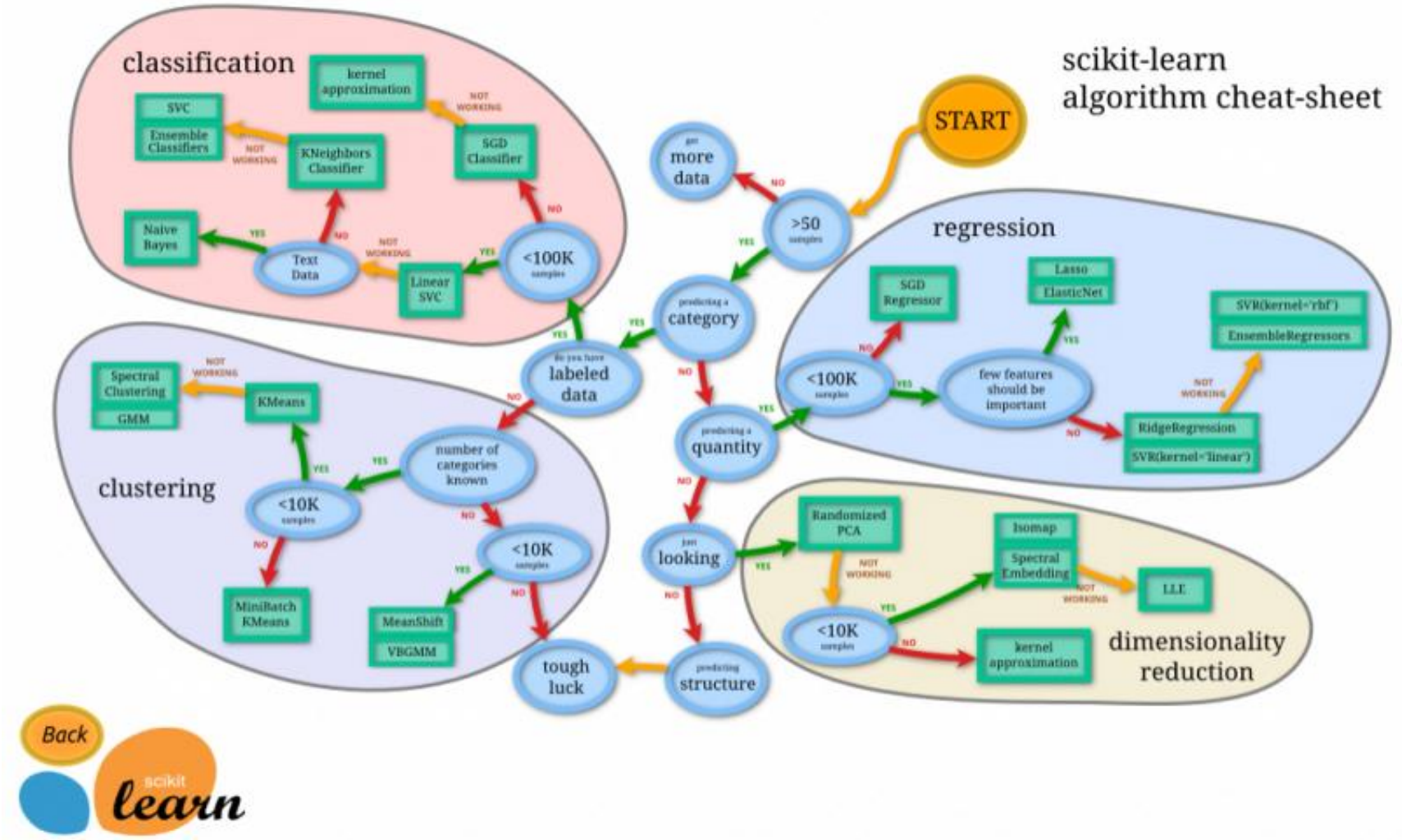
様々な分析モデルが利用される

- 教師あり学習による予測モデル
- 因果推論モデル
- 行列分解などの次元縮約
- トピックモデルによるクラスタリング
- Embedding（埋め込み）表現モデル
- ネットワーク分析　　etc.

使いこなすために必要なスキルは？

scikit-learnで可能な分析

scikit-learn：Pythonの機械学習ライブラリ



出展：https://techacademy.jp/magazine/17375

データサイエンティストに必要なスキル

ビジネススキル

- ビジネスに対する理解
- ロジカルシンキング
- ビジネス課題の設定
- 分析結果の活用力

ITスキル

- プログラミングスキル
- 広いIT知識
- データエンジニアリング力
- データベースの知識

統計解析スキル

- 統計学
- 数学
- データ分析手法の理論
- データ分析ツールの利用スキル

どちらかと言うと、ITスキルと統計スキルよりも、ビジネススキルが差別化のポイントになりつつある

事例紹介

ビジネスデータ分析の取り組み事例

- ECサイトにおける施策（リアルタイムクーポン）効果の向上を目的としたマルコフ潜在クラスモデル分析
- 気象情報とTweetデータの統合的分析による体感気温の定量化とその需要予測への利用
- 購買履歴に基づくポイントカードユーザのクレジット切り替え分析モデル
- 中古ファッションアイテムの販売価格推定モデル、価格維持期間の適正化モデル
- 社内チャットアプリ上の会話履歴のネットワーク分析、テキストデータ分析によるコミュニケーションタイプ分析
- 生花ECサイトにおける顧客の閲覧履歴・購入履歴データの統合分析モデル
- プラットフォームビジネスの利用履歴データに基づく顧客生涯価値評価モデルの構築

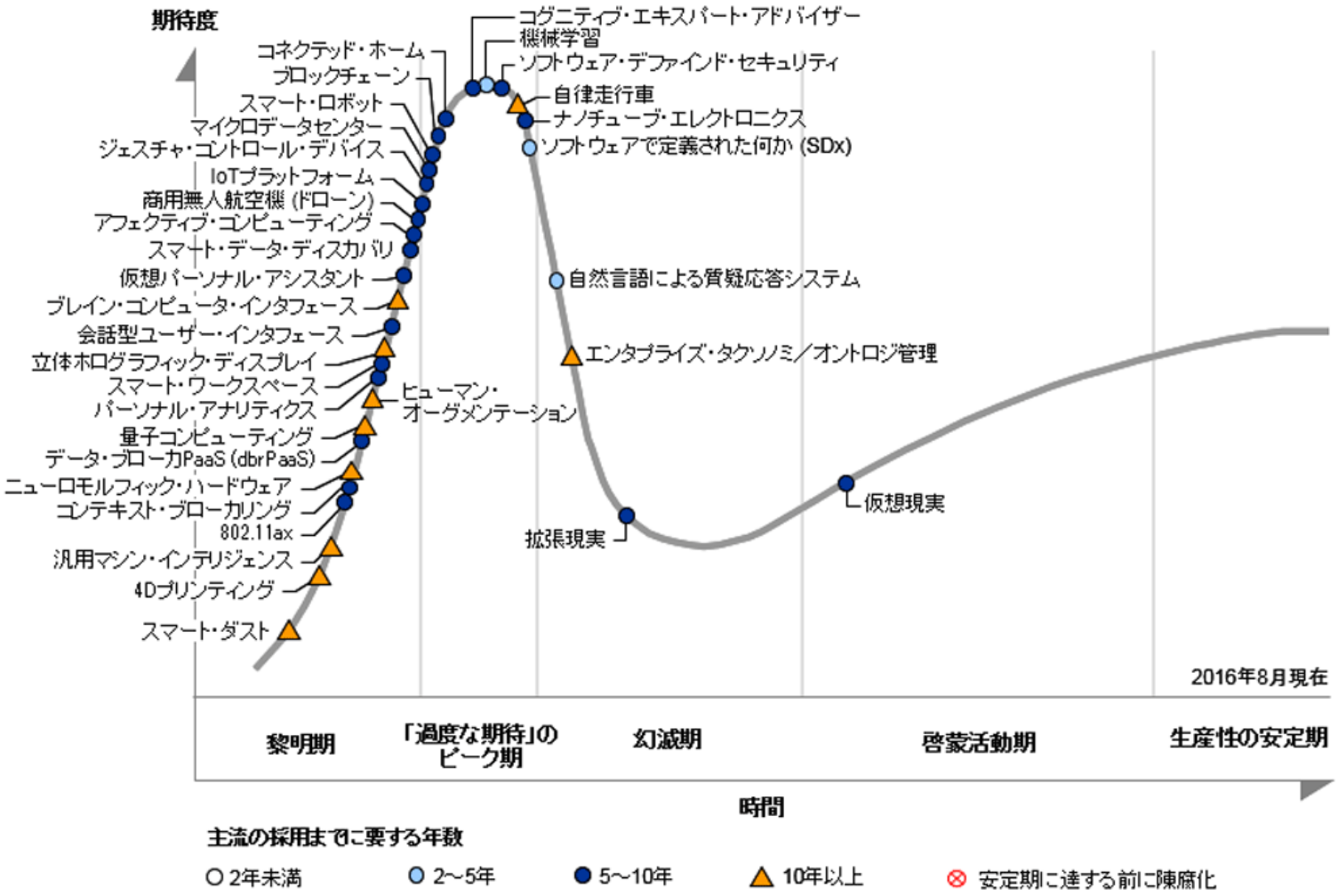
今後の展望

おわりに

ビジネスにおけるデータ活用は、囲碁や将棋のようなルールの決まったゲームとは異なる。ビジネスの構造やルールは絶えず変化し、対戦相手も時として明確ではない。

- 「機械学習」、「自然言語処理」などは、これから幻滅期に入る。この後の啓蒙活動期へ向けた地道な取り組みが必要
- 産業界の様々な事例、データの分析経験を積み上げ、ビジネスアナリティクスのノウハウを蓄積する必要がある。
- 産学の連携は大変重要

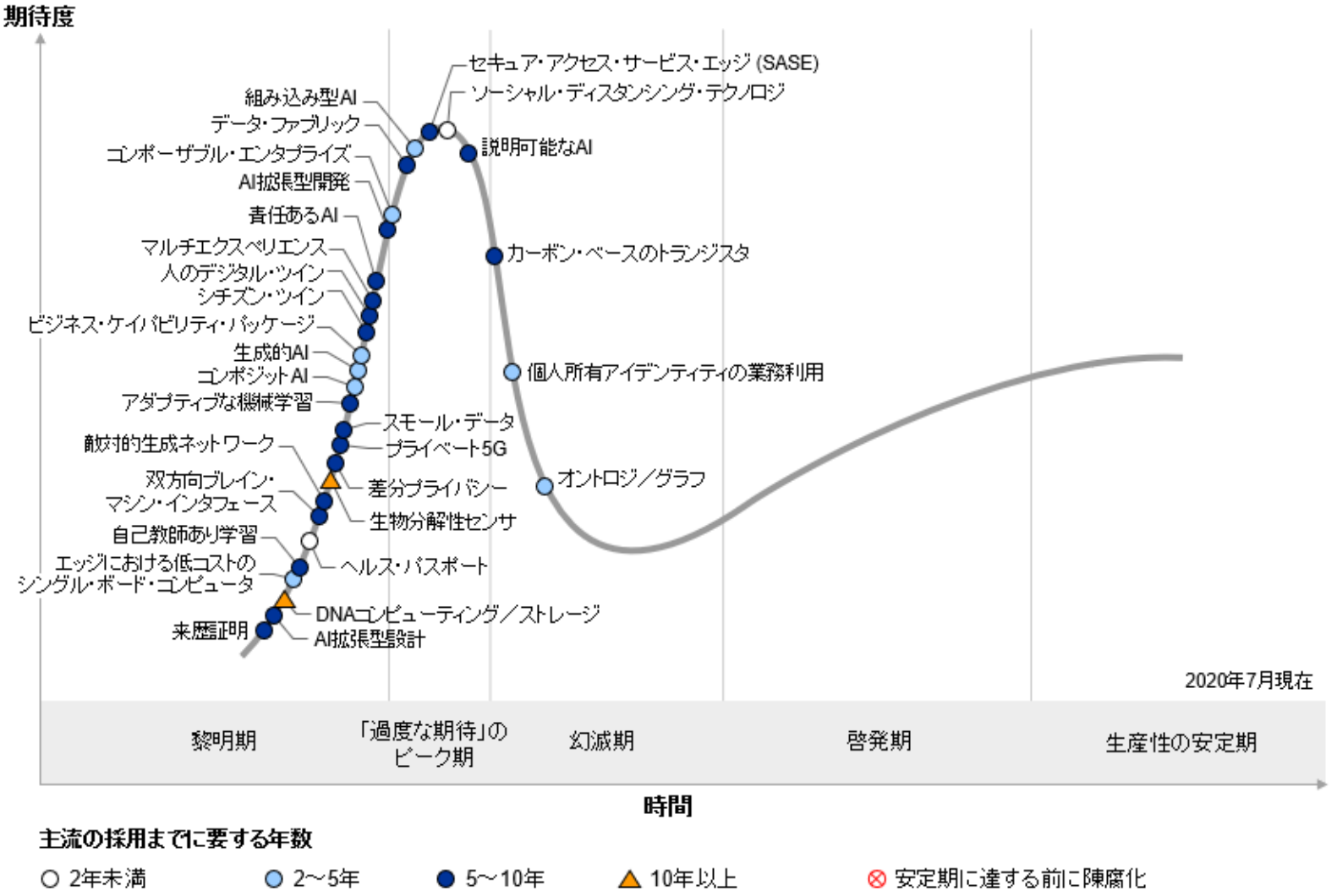
ガートナー社によるパイプサイクル（2016年）



出典：ガートナー (2016年8月)

https://www.gartner.com/jp/promo/hype-cycle-report-list　(2016.10.アクセス)

ガートナー社によるパイプサイクル（2020年）



出典：ガートナー (2020年9月)

https://www.gartner.com/jp/newsroom/press-releases/pr-20200819

https://www.gartner.com/jp/promo/hype-cycle-report-list

おわりに

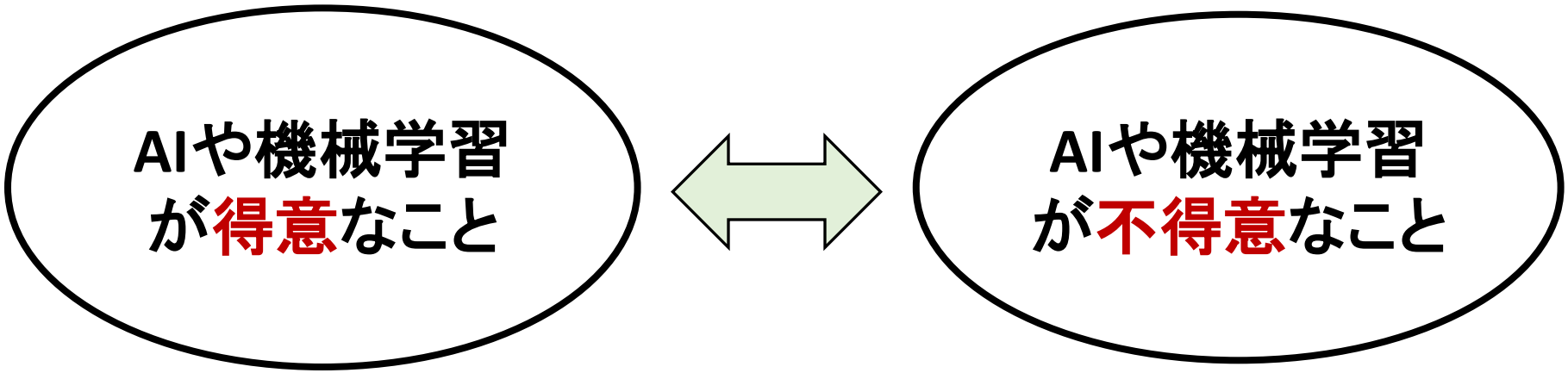
- 人工知能、ビッグデータ、機械学習は、期待値が高騰気味。今後、試練のときが訪れる。
- 一方、ビジネス分野におけるデータ活用の重要性は失われない。
- 様々な新しい分析技術のビジネスアナリティクスにおける有用性について、多くの適用事例を通じて、評価を定めていく必要がある。
- IoT の流れは、オープン化の流れ（？）
- 次に来たる IoT時代へのパラダイムシフトに備える

人間とAIがパートナーとなって仕事をする時代へ

おわりに

私が考えるキーポイント

データ分析者がAIや機械学習を使いこなすために



ご清聴ありがとうございました。