資料２

AI活用と果たすべき責任

～データの来歴とAI～

日本マイクロソフト株式会社 業務執行役員 NTO

田丸 健三郎

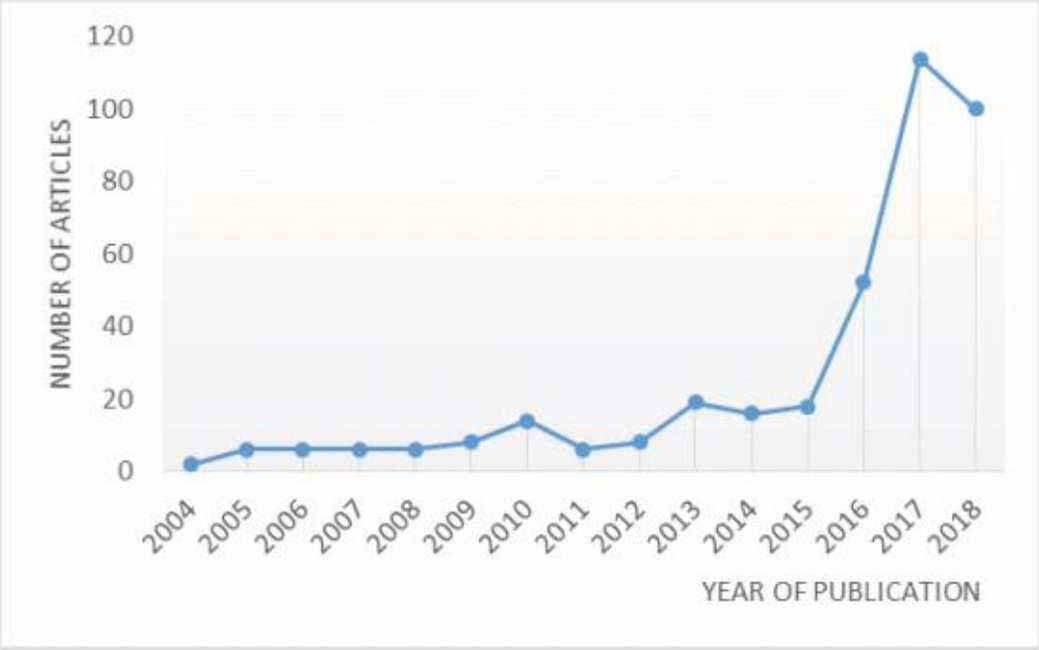
ソフトウェア・プログラムと製造物責任

（製造物責任法第二条）　この法律において「製造物」とは、製造又は加工された動産をいう。

消費者庁「逐条解説 製造物責任法」

ソフトウェア自体については、無体物であり、製造物責任の対象とはし ていない。ただし、ソフトウェアを組み込んだ製造物については、本法の対象 と解される場合がありうる。ソフトウェアの不具合が原因でソフトウェアを組 み込んだ製造物による事故が発生した場合、ソフトウェアの不具合が当該製造 物自体の欠陥と解されることがあり、この場合、その欠陥と損害との間に因果 関係が認められるときには、当該製造物の製造業者に本法に基づく損害賠償責 任が生ずる

“Right to Explanation” – GDPRへの備え



Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)

European Union regulations on algorithmic decision-making and a "right to explanation"

Bryce Goodman, Seth Flaxman

(Submitted on 28 Jun 2016 (v1), last revised 31 Aug 2016 (this version, v3))

We summarize the potential impact that the European Union's new General Data Protection Regulation will have on the routine use of machine learning algorithms. Slated to take effect as law across the EU in 2018, it will restrict automated individual decision-making (that is, algorithms that make decisions based on user-level predictors) which "significantly affect" users. The law will also effectively create a "right to explanation," whereby a user can ask for an explanation of an algorithmic decision that was made about them. We argue that while this law will pose large challenges for industry, it highlights opportunities for computer scientists to take the lead in designing algorithms and

IGF 2018 WS #421 Algorithmic transparency and the right to explanation

Theme:
Cybersecurity, Trust and Privacy

Subtheme(s):
ALGORITHMS

Organizer 1: Civil Society, African Group
Organizer 2: Civil Society, Western European and Others Group (WEOG)

Format:
Break-out Group Discussions - 60 Min

Interventions:

- Alex Comninos and Deborah brown (APC) will be moderators - Jelena Jovanovic (cyber security professional) will provide an overview of the concepts of algorithmic transparency, algorithmic justice, algorithmic bias and real life examples of the effects of algorithms from an informatic security perspective. - Vidushi Marda (Article 19) will provide an overview of the human rights aspects of automated decision making. She will focus on the GDPR Article 22 and the EU guidelines on Automated decision-making. She will provide a policy and human rights perspective.

The ICANN 2018 scientific programme includes the following workshops.

Interpretable Methods for Machine and Deep Learning

Description: Machine learning-based methods and, in particular, deep neural networ have proven extremely successful in a wide variety of tasks but remain for the most part extremely opaque in terms of what they learn and how they use their acquired knowledg to make predictions. More specifically, without any insight into the network, the task of determining whether the system has properly integrated a specific concept is very difficult and the subsequent validation of the method for critical activities (e.g. driving, medicine,

2018 Workshop on Human Interpretability in Machine Learning (WHI)

July 14, 2018

Stockholm, Sweden

Overview

The Third Annual Workshop on Human Interpretability in Machine Learning (WHI 2018), held in conjunction with ICML 2018 and the Federated Artificial Intelligence Meeting, will bring together researchers who study the interpretability of predictive models, develop interpretable machine learning algorithms, and develop methodology to interpret black-box machine learning models. They will exchange ideas on these and allied topics, including:

- Quantifying and axiomatizing interpretability,
- Psychology of human concept learning,
- Rule learning,
- Symbolic regression,
- Case-based reasoning,
- Generalized additive models,
- Interpretation of black-box models (including deep neural networks),
- Causality of predictive models,
- Visual analytics, and
- Interpretability in reinforcement learning.

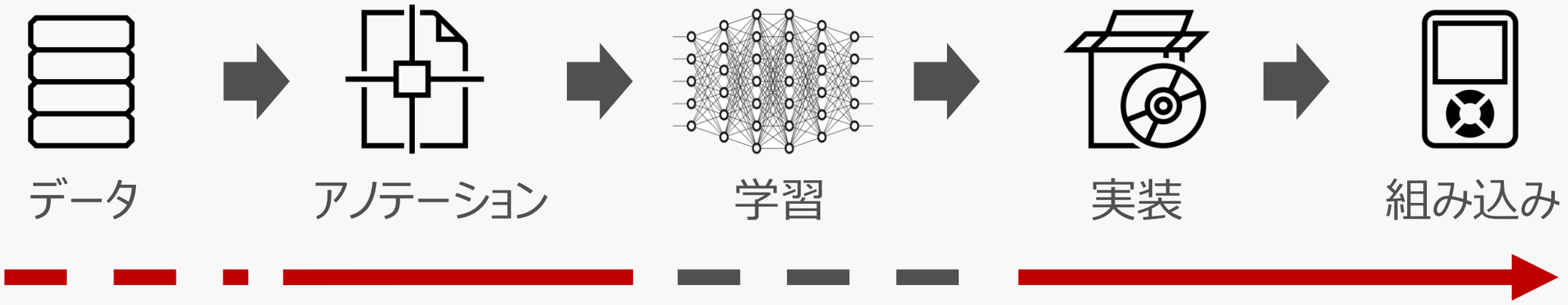
検討されている主な説明手法

- 予測・認識とデータの特徴
  - データの分布（特徴）と結果を用いて説明
- 予測・認識と学習データ
  - データの有無、偏りを操作することによる結果への影響と予測
- インファレンスプロセスの可視化

帰納的推論ではデータは重要な説明要素の１つ

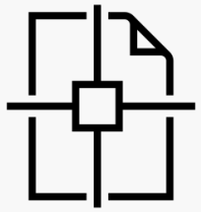
- データソース、ボリューム、分布、クレンジング、アノテーション

AIとアルゴリズム実装と想定される責任範囲



そもそも信頼できるデータとは何か

観光、健康、電力、河川、農業、交通・物流、ソーシャルなどIoTデバイス、人、**AI** に至るまで非常に多くのデータソースが存在する。一方で認証されていないデータソース、基準が明確でないデータが多く使用されている。

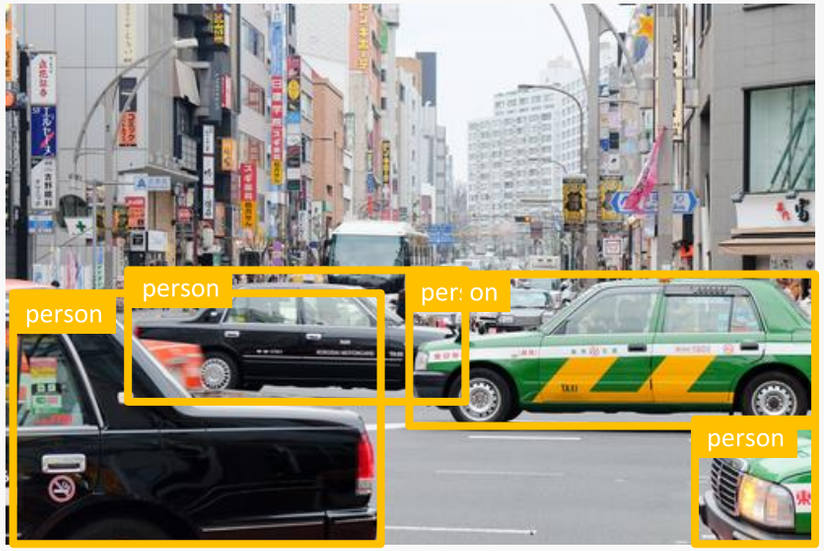


学習データの重要な要素であるクレンジング、アノテーション

- 作業者、品質基準、品質保証・・・規格

想定すべき悪意あるアノテーション

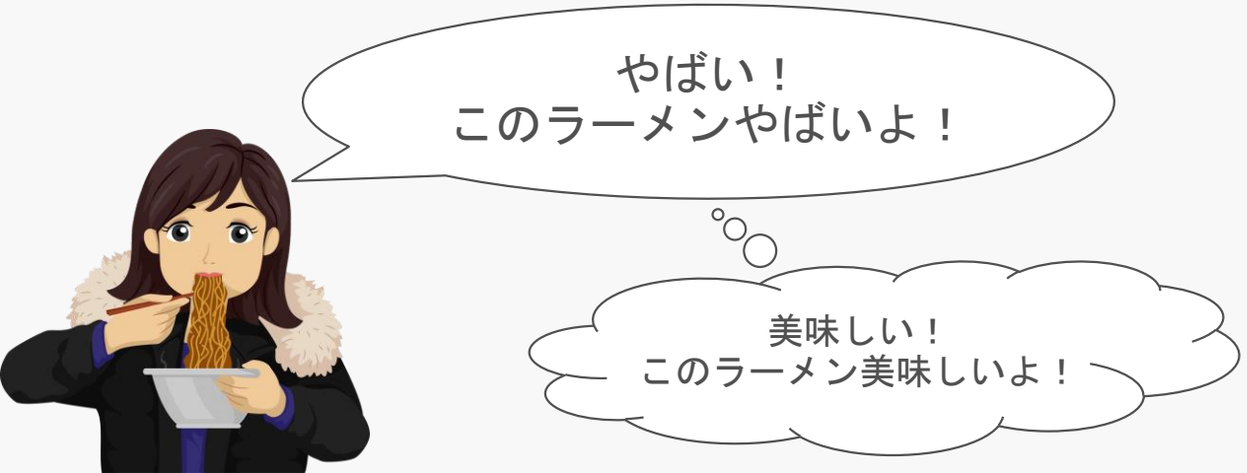
- 様々なアノテーションの品質と基準
- 「誤り」と「悪意」

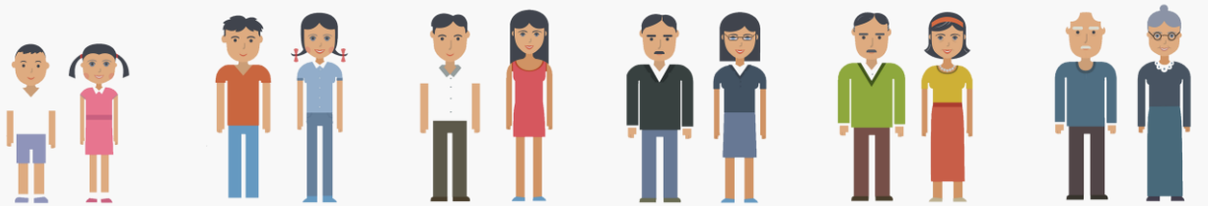


重要性を増す複数のモダリティによるモデル構築

- マルチモーダル深層学習で重要となる複数の異なるデータ
- ヘテロジニアスなデータによる複合モデルの構築
- 継続的な更新が必要なデータ

テキストデータ、音声データだけでは「やばい！」の意味を評価できない。行動データとの組み合わせで初めて「やばい！」の評価ができる。

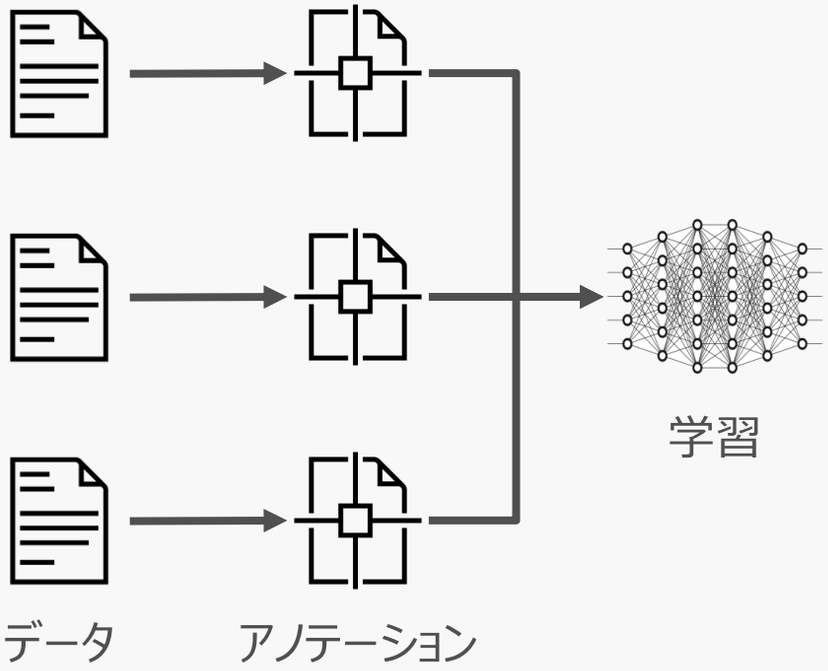


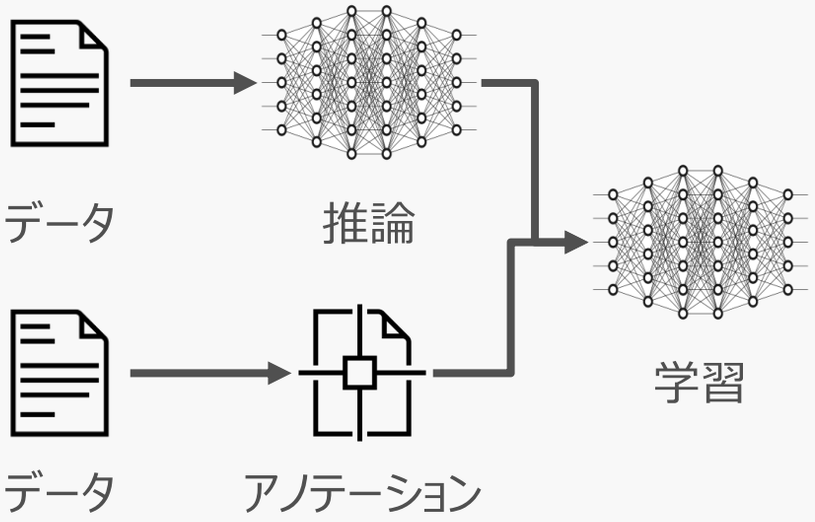


変化し続ける言葉。10年前の10代と現在の10代では使用する言葉が大きく異なる。AIモデル品質の維持、向上には継続的なデータ収集、更新が欠かせない。

想定が難しいデータソース

- データソース
  - センサー（デバイス）、統計データ、サービス（推論結果）
  - そもそも、データ生成源を常に確認、もしくは保証することは可能なのか？





有機的に相互作用するAI/IoT社会

保守計画
車両情報
故障予知
混雑予測
駅員・警備計画
△△線運行情報
○○線運行情報
イベント開催情報
道路工事情報
交通渋滞予測
TV番組情報
物流
集荷予測
売上予測
販売計画
混雑予測
天候情報
空調負荷予測
空調制御計画

循環の問題

- 不透明化するデータの確かさ
- 影響を受ける動態、揺らぐモデル

ハイマン・ミンスキー
『金融不安定性仮説』

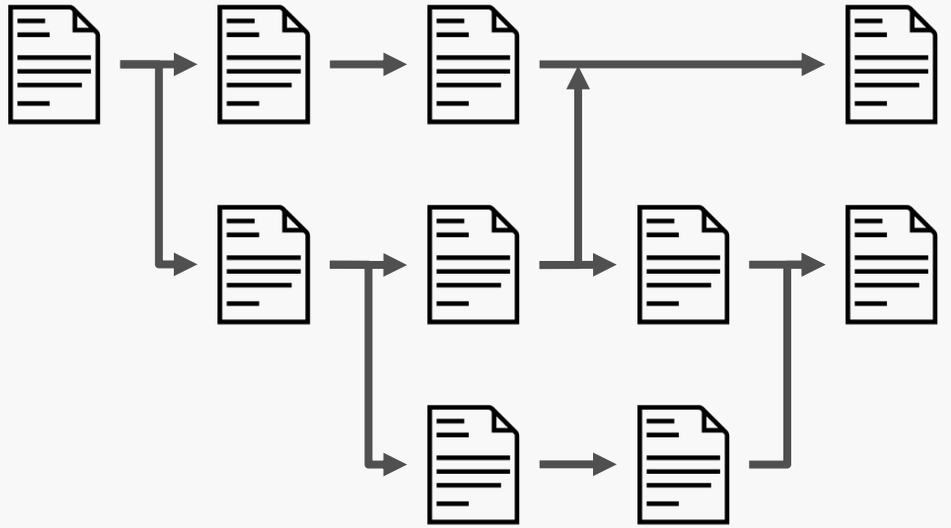
ポール・マカリー「人間はリスクを取ることで利益を得る期間が長くなればなるほど、リスク・テークに対して無警戒になる。人々がリスク・テイクに無警戒になると、自己実現的に相場が上昇し、すべての人が同時にリスク志向を高めると、それによってリスク・プレミアムが低下する。これが担保の価値を押し上げ、さらに多くのレバレッジの利用が可能になり、ゲームは続いていく。人間は本質的に正循環的である・・・それこそがミンスキー理論の核心である。」

AIにおけるデータとはリスク要因であり、無意識によるリスク・テークである。

来歴 [Provenance]

- 【何が、誰が】何時、どのように生成したデータなのか
- 【何が、誰が】何時、どのようにクレンジングしたデータなのか
- 【何が、誰が】何時、どのようにアノテーションしたデータのか
- 【何が、誰が】何時、どのよう評価したデータなのか

- 重要となる基準とプロセスの標準化



世界をリードする取り組みの例

AI Data Consortium
AIデータ活用コンソーシアム

一般社団法人 AIデータ活用コンソーシアム

- AIの為の円滑なデータ流通の実現を目指し、2019年3月に設立
- 主な活動
  - 複雑な商流に対応した知財・契約の検討とスマートコントラクトの実装
  - **データの来歴管理手法の確立、標準化および実装**
  - **データ、アノテーション評価基準の標準化と実装**
  - データ流通を加速する基盤の実現とサービスの提供（OSS公開含む）

代表理事・会長
- 長尾 真
  元京都大学総長 / 元独立行政法人情報通信研究機構理事長 / 元国立国会図書館長 / 前国際高等研究所所長

理事・副会長
- 渡部 俊也（知財・契約担当）
  東京大学 政策ビジョン研究センター 教授/副センター長大学執行役・副学長、工学系研究科技術経営戦略学専攻 教授（兼務）
- 杉山 将（AI研究担当）
  理化学研究所 革新知能統合研究センター センター長、東京大学 大学院新領域創成科学研究科 教授
- 井佐原 均（データ収集・活用担当）
  豊橋技術科学大学情報メディア基盤センター センター長・教授
- 田丸 健三郎（データ基盤担当）
  日本マイクロソフト株式会社 業務執行役員 NTO