

# 「イントロ・傾向分析」より



## ヤン ジャクリン

北京生まれ・米国籍・日本育ち（茨城県民8年, 東京都民15年）

日本語・英語・中国語のトライリンガル

私の好きなものは：

言語（日・英・中 80%, スペイン語5%, 仏語5%, [Python 10%](#)）

コミュニケーション全般、文章の読み書き、運動

### 【経歴】

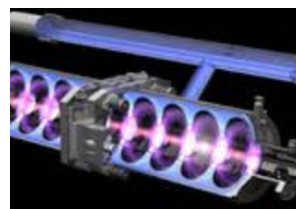
2010.3 東京大学理学部卒業

2015.3 東京大学理学系研究科 物理学専攻 博士課程修了（理学博士）

～2017.6 高エネルギー加速器研究機構（茨城県つくば）

素粒子原子核研究所 博士研究員

- ・ 主要研究：加速器・ナノビーム・検出器・ヒッグス粒子・暗黒物質
- ・ 各地大学で物理学の非常勤講師、産学連携分野の広報・通訳



2017.7～ 株式会社GRI（現職）

データ分析官・講師

最近の主な活動：

- ・ マーケティング分野の分析、データ整形・可視化分析
- ・ データサイエンス分野の講座、法人研修を提供
- ・ データサイエンス人材育成カリキュラムの設計



- 一般社団法人日本ディープラーニング協会（JDLA）が実施  
<https://www.jdla.org/business/certificate/>
- 目的  
**ディープラーニングを事業に活かすための知識を検定する**
- 試験概要：
  - **多肢選択式の知識問題 220問程度**（シラバスより出題）
  - **人工知能やディープラーニングに関する幅広い内容が出題される**
- 試験時間：**120分**
- 試験形態：**オンライン実施**（自宅受験可）
- 受験料：一般 12,000円（税抜） 学生 5,000円（税抜）
- 試験日：年3回、日程確認と申込は <https://www.jdla-exam.org/d/>

## ■ Part1 人工知能（AI）

人工知能の定義や歴史、議論されている問題などに出てくる重要なワード

## ■ Part2 機械学習

機械学習の基本概念、具体的な手法の名称、特徴、用途、精度評価の指標

## ■ Part3 ディープラーニングの概要と手法

ニューラルネットワークの仕組み、学習法と精度向上のための工夫、計算リソース、ディープラーニングの様々なテクニックと応用

## ■ Part4 ディープラーニングの研究分野と発展可能性

画像認識、自然言語処理、音声処理、ロボティックス、マルチモーダルなど、ディープラーニングの最新動向

## ■ Part5 ディープラーニングの応用に伴う社会的課題

現在議論されている法律や倫理についての広い知識が問われる

- G検定には難易度の高い出題もみられるが、必ずしも受験の際に全てを知っていることが期待されているわけではない
- 「受験を機に、解けなかった問題については自主的に調べて知識を広げてください」というメッセージのように受け止められる  
(個人の推測)

合格はゴールではなく、ジェネラリストとしてのスタート

継続的に勉強を続ける姿勢、それこそがジェネラリストに求められること！

問題：第2次AIブームの特徴として、適切なものを1つ選べ

Part1より

1. 推論と探索を可能とするAI
2. 大量な知識を蓄え、専門家のように返答できるAI
3. AIによって、特徴表現を自動的に抽出可能になった
4. AIが迷路やチェスなど明確なルールのある問題に対応可能になった

問題：過学習の原因の候補として、適切ではないものを1つ選べ

Part2より

1. モデルの関数が高次元で、複雑な形をしている
2. 学習データの量やバリエーションが多すぎる
3. 特徴量の数が非常に多い
4. お互いに相関が強い特徴量が一定数ある

問題：（ア）に当てはまるものとして、最も適切なものを1つ選べ

Part3より

再帰型ニューラルネットワーク（RNN）は（ア）ことが可能であることが特徴の1つである。従来のRNNが抱える困難の1つとして（ウ）が挙げられる。現在RNNは（イ）などに活用されている。

1. 重要度の高い情報のみ残し、情報を効率よく圧縮する
2. 過去の入力を現在の出力に反映させる
3. 教師データを含まない学習データでも、分類に使用できる
4. 試行錯誤を通じて、報酬が最大となるような行動を見出す

## 「Part1 人工知能とは」より

一般的な定義：

**「人間が持つ知的な情報処理能力を機械に持たせること」**

**「周囲状況（入力）によって行動（出力）を変える能力を持つ機械」**

- **専門家の中でも解釈が様々で、いまだに定まっていない**  
自動化、推論、探索、認識, …どこまでがAIの条件？

■ AIを人が行う**知的作業の代替装置**と考えることが多い

目	⇒	画像認識
手足	⇒	ロボット制御
耳	⇒	音声認識
計算	⇒	数値の予測
会話	⇒	文章の作成



# 人工知能

## ルールベース

ルールに従って判断する

条件Aの下で入力B ➡ 出力C

## 機械学習

学習データから自動的に情報を整理し、裏の法則を見つける

データ				正解



## シンギュラリティ (Singularity)

### 技術的特異点

AIが自ら人類の知能をこえる知能を生み出せるようになる転換点

提唱者：

- 数学者**ヴァーナー・ヴィンジ**

「30年以内に技術的に人間を超える知能がつくられる」

- **レイ・カーツワイル**博士

「2029年にAIが人間並みの知能を備え、2045年に技術的特異点がある」

この「**2045年問題**」の根拠となっている理論が「**収穫加速の法則**」

- 「技術進歩の性能が直線的ではなく指数関数的に向上する」
- 一度技術が進歩するとその技術が次の進歩までの期間を短縮させる

**「Part2・機械学習」 「Part3・ディープラーニング」より**

次に、AI・機械学習の実装にとって、重要な概念である  
「**特徴量エンジニアリング**」について説明する

【キーワード】

**特徴量、特徴量設計、特徴量エンジニアリング**

**構造化データ、非構造化データ**

**データの前処理**

**カテゴリカル変数**

**ラベル・エンコーディング、ダミー変数**

**One-hot エンコーディング**

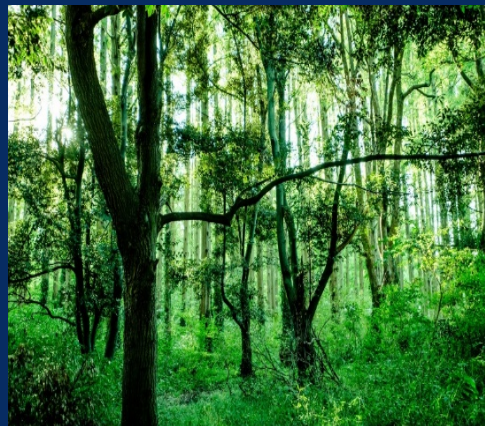
**欠損値**

# アンサンブル学習器

複数の単純なモデル（弱学習器）を組み合わせることで、過学習を抑制し精度を改善する技術



単純な決定木



アンサンブル学習器  
「3人よれば文殊の知恵」

- **バギング**

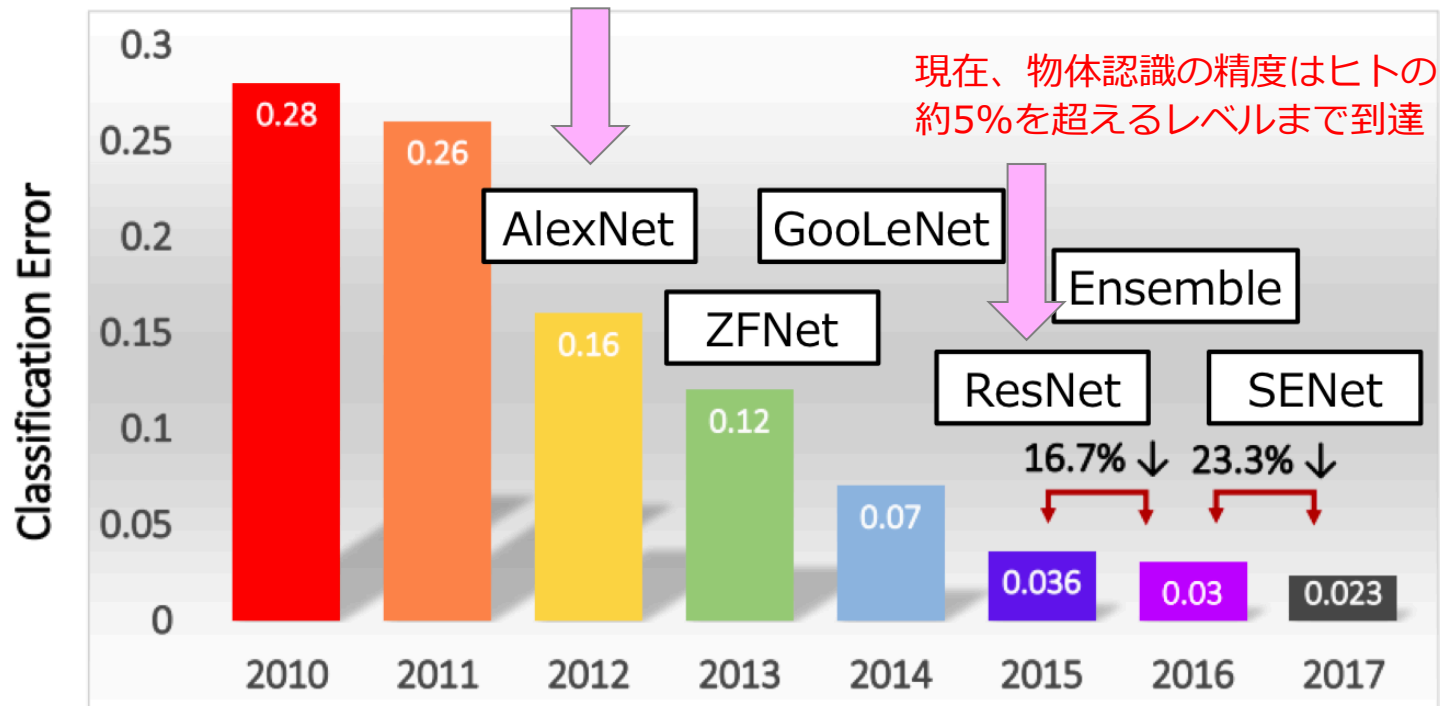
毎回一部のデータを抽出して少しずつ異なるモデルを並列に多数作り、最後に統合することで、個々の弱学習器のバイアスを相殺（例:ランダムフォレスト）

- **ブースティング**

弱学習器を1つずつ順に構築し、新しい弱学習器を構築する時に以前構築された弱学習器の結果を活用（例:勾配ブースティング木）

- ディープラーニングが理論として初登場した1980年代以来、データ量の増加と計算リソースの改善により進歩してきた
- メディアに大きく注目されるきっかけは、2012年に開催された「ILSVRC2012」（画像に写る物体のラベルを予測するコンペ）

ディープラーニングを利用したモデルが、  
前年までの誤差率を10%以上改善



[http://image-net.org/challenges/talks\\_2017/ILSVRC2017\\_overview.pdf](http://image-net.org/challenges/talks_2017/ILSVRC2017_overview.pdf)

## ■ 簡単に言うと？

### 行動を学習する機械学習手法

「ある状態下で、最大限のご褒美をもらうために、どんな行動をとるべきか」

## ■ 特徴

- 試行錯誤や探索を通じて意思決定のルールを導き出す手法
- 教師ラベルがなくても学習できる（ここが他の機械学習と違う）

## ■ 応用例

- 自動運転
- ルールが複雑なゲーム（将棋、囲碁; AlphaGoなど）
- ロボティクス



## 「Part5 社会への応用・法律・倫理」より



G検定では **AI関連分野の法律・規則についても出題される**  
試験対策で軽視されやすく、**合否の決め手になることも！**  
AI分野の仕事に携わる全員が知っておく知識が多い

<2020年3月の試験後、Twitterにおける最頻出用語>

## Part5で出題されやすい用語例

- 知的財産法
- 特許発明の要件
- 著作権
- 不正競争防止法
- 限定提供データ
- 個人情報保護法
- 匿名加工情報

キーワード	登場回数
個人情報（保護）	36
法律・倫理	28
営業秘密	25
XAI	12
音声認識	6
時事問題	3
強化学習	3
統計学	2

試験日（2020年11月7日）から1か月間にわたるtwitterの投稿を「G検定」で検索をかけてデータを収集

- 営業秘密とは以下の要件を満たすもの（不正競争2 VI）

### ① 秘密として管理されていること（秘密管理性）

従業員等からみて、その情報が秘密としたい情報であることが分かる程度に、アクセス制限や<sup>秘</sup>表示などの秘密管理措置がなされていること

### ② 有用な技術上または営業上の情報であること（有用性）

ネガティブ・インフォメーション（例：失敗した実験のデータ）にも有用性が認められることがある

### ③ 公然と知られていないこと（非公知性）

保有者の管理下以外では一般的に入手できない

- 営業秘密の侵害に対する措置：

- ・ 民事的救済：差止請求権、損害賠償請求権、信用回復措置
- ・ 営業秘密侵害行為等については刑事罰が科せられることもある

## 問 1

(ア) に最もよく当てはまる選択肢を 1 つ選べ。

リカレントネットワーク (RNN) は内部に (ア) を持っているニューラルネットワークである。RNN の特徴の 1 つは (イ) である。画像認識に使われるニューラルネットワークと違って、(ウ) を入力データとして扱うことができる。RNN は (エ) であるため、ニューラルネットワークの中でも特に (オ) が起きやすいという難点を持っている。

1. プーリング層
2. エンコーダ・デコーダ
3. スキップグラム
4. フィードバック機構

## 問 26

以下の文章を読み、最も適切な用語の組み合わせの選択肢を 1 つ選べ。

ニューラルネットワークの学習において、具体的には、(ア) の値を (イ) にするパラメータを見つけることを目的としている。

1. (ア) 誤差関数 (イ) 最大
2. (ア) 損失関数 (イ) 最小
3. (ア) 活性化関数 (イ) 最大
4. (ア) シグモイド関数 (イ) 最小