

2022年6月16日(木) 15:40~15:58

説明可能AI(XAI)・共進化AI(CAI) 基盤技術の開発と産業応用

横浜国立大学 大学院環境情報研究院 教授
YNU人工知能研究拠点長・(株)マシンインテリジェンス取締役CTO

長尾 智晴

電子メール : e-mail:	nagao@ynu.ac.jp
長尾研究室URL:	https://nagao-lab.ynu.ac.jp/
YNU人工知能研究拠点:	https://ai.ynu.ac.jp/
(株)マシンインテリジェンス:	http://machine-intelligence.co.jp/

発表内容

- 説明可能AI (**XAI**)・共進化AI (**CAI**)
XAI: eXplainable AI, **CAI**: Co-evolutional AI
- 本研究（「**進化的機械知能に基づくXAIの基盤技術と産業応用基盤の開発**」）の概要
- 本研究の成果例
- 成果の産業応用について



説明可能AI (XAI) ・ 共進化AI (CAI)

● AI ・ 機械学習の発展の推移

- 第1世代：論理・ロジックに基づく推論
→ 論理的な思考を機械に代替させる。
- 第2世代：人の知識に基づく処理 → 知識ベースの重要性。
- 第3世代：機械が人のように学習 → 高精度な判定。説明性は軽視。



説明可能AI (XAI)

機械 → 人

→ 機械による処理・判定根拠を可視化して人に説明することで**人が機械の考えを理解できる。**

人がもつ知識を機械に与えることが考えられる

知識・知能の相互のやりとり・コミュニケーション・コラボが可能になる。

共進化AI (CAI)

機械 ↔ 人

→ **人と機械学習が相互に知能を高め合える。**

機械学習が現在の**事例ベース学習**(IBL:Instance Based Learning)から脱却して、**説明ベース学習**(EBL:Explanation Based Learning)へと変わることができる。

深層学習の特徴とその改善のための最適化

● 深層学習の長所

- 大量の学習用データがあれば **高精度**の **入力** → **出力**(分類・回帰)が可能.
- 人が入力の**特徴**や**処理アルゴリズム**を考える必要がない.
- **ライブラリ**・**開発環境**などが充実している.

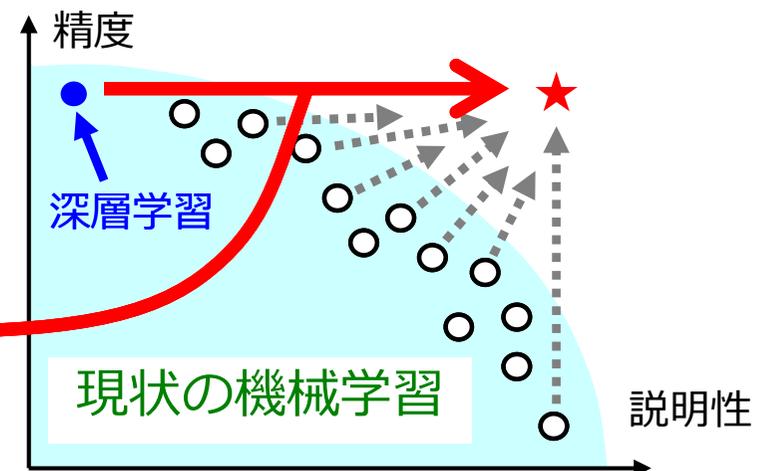
● 深層学習の短所

- **複雑な回路**で判断根拠・処理手順(機序)が分りづらく **説明性が著しく低い**.
- 決定すべき結合荷重などが多いため **大量の学習データが必要**.
- **人の知識が組み込みづらい**.
- 説明責任を果たせないため **企業で利用しづらい**.

- ◆ **説明性が高く, 少量の学習データでも学習でき, 人の知識の組み込みや対話**
がし易く, 産業応用し易いように
深層学習を最適化・改良する必要がある.

本研究

「**進化的機械知能**に基づくXAIの基盤技術と産業応用基盤の開発」



本研究の概要

研究代表者：長尾智晴（横浜国立大学）

題目

進化的機械知能に基づくXAIの基盤技術と産業応用基盤の開発

課題

深層学習などの機械学習の判断根拠や処理内容が人にとって分かりづらく、人の知識を利用するなど、**人と機械が共進化するようなAI基盤技術・産業応用が存在しない。**

狙い

説明できるAIを発展させた**共進化AI「進化的機械知能」の基盤技術を開発し、産業応用基盤を開発**してヘルスケア・製造・サービス分野などで**産業応用**することを旨とする。

研究開発内容

3) 説明できるAIの産業応用（横浜国立大学・東京医科大学・キューピー(株)）

- ⑦医療を対象とした産業応用：マイクロRNAによるがんリスク判定・健康社会実現
- ⑧医療・ヘルスケア以外の産業応用：XAI経営判断支援・業務支援、新規事業創出
- ⑨産業応用技術・基盤を用いた事業創出：XAI導入支援を行う新規事業創出

説明できる
AI技術の
産業応用

2) 産業応用技術・基盤の開発（横浜国立大学ほか）

- ④産業応用技術・基盤の定義：ヒヤリング・仕様策定など
- ⑤産業応用技術・基盤プロトタイプ構築：ソフト試作
- ⑥産業応用技術・基盤構築：API・ライブラリ・DB等

産業応用の
結果に基づく
技術改良
/新規研究
創出

1) 説明できるAI基礎技術“進化的機械知能”の開発（横浜国立大学）

- ①機械学習の説明性・精度向上：深層学習・非深層学習の機械学習の最適化
- ②人の知識を利用する説明の実現：定義型(what)知識のMLへの導入と相互作用
- ③説明(解法)の半自動生成の実現：手続き型(how)知識のMLへの導入と相互作用

ベースとなる技術：進化的機械知能

- 進化計算法などの最適化技術により深層学習などの機械学習の最適化・全自動構を目指す独自技術。進化計算と深層学習を熟知している研究者でなければ実施不可能。

横浜国立大学・長尾 (研究代表者) & 白川 (研究分担者) :

教授 長尾 智晴 : 1990年初頭から進化計算法の研究に従事。画像処理・ニューラルネットや機械学習の最適化を中心に多くの実績あり。産学連携にも注力。これまでに150社を越える企業とコラボあり。進化的画像処理のAIベンチャーを起業して取締役CTOを兼務中。

准教授 白川 真一 : 長尾研出身。進化的機械学習の研究に従事。深層回路の構造最適化の理論・手法などを研究開発。AutoMLなどで多数の実績あり。バイオ他異分野との産学連携にも注力。

本研究「進化的機械知能に基づくXAIの基盤技術と産業応用基盤の開発」



※“進化的画像処理”は横浜国立大学の登録商標です。

進化的画像処理®ソフトウェア CRAFT-IT (1本500万円) :

横浜国立大学発ベンチャー(株)マシンインテリジェンス (2008年発売), 開発者 白川真一ほか, 取締役CTO 長尾智晴 など

複数の商用ソフトの開発実績あり。

進化計算法・機械学習の最適化・自動化などに関して300本以上の論文・国際会議・書籍・特許などの実績あり。

産業応用に必須な複数の特許

本事業に関するものだけで
8件(出願済)+10件(取得済み)

特願2019-137811,
特願2021-032097,
特願2021-041855,
特願2020-75691,
特願2020-75692,
国際出願番号PCT/JP2018/028633 ほか

他者に対する差別化・権利化のための知財の確保

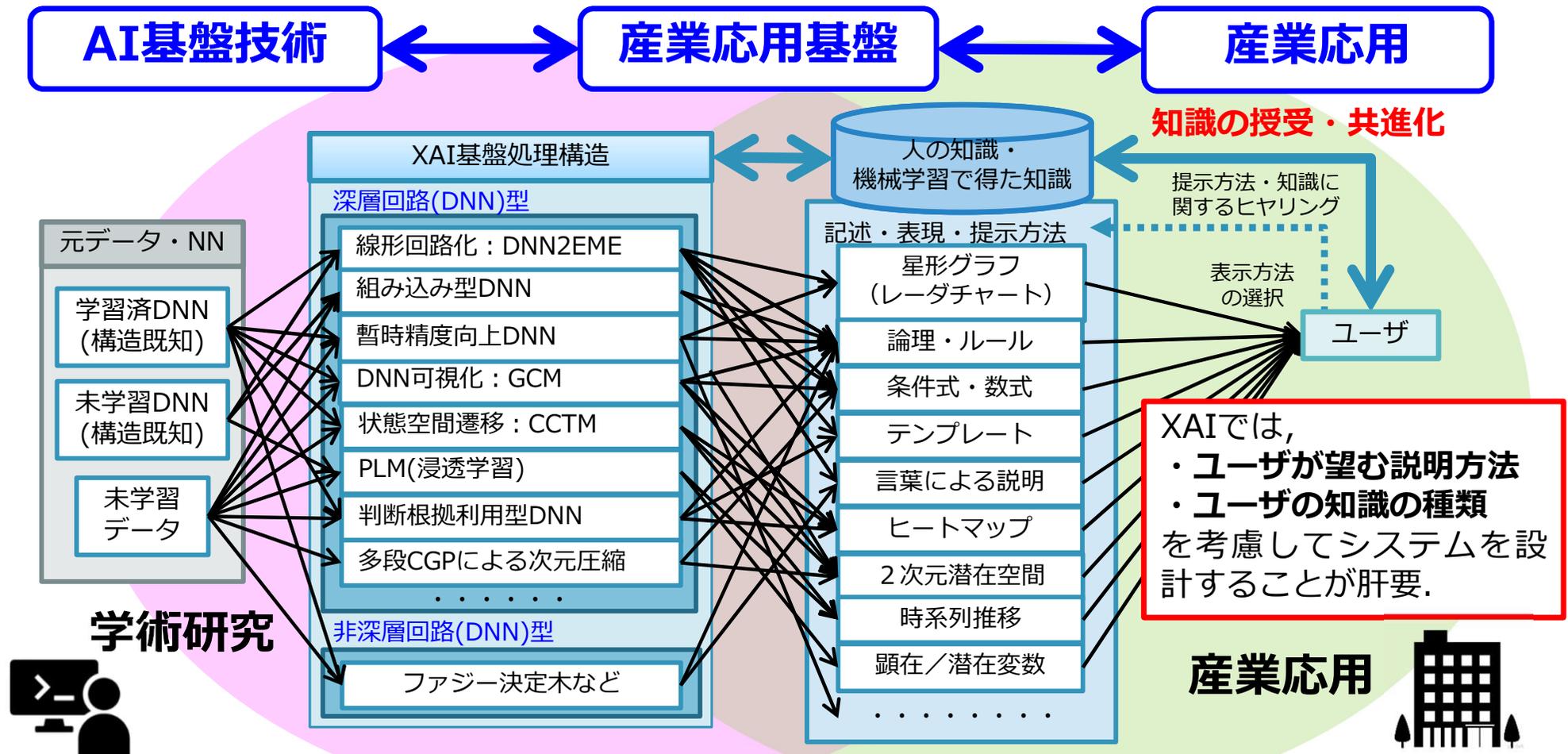
長尾智晴 他 : 遺伝的手法による神経回路網の構造進化 ; 電子情報通信学会論文誌, Vol.J75-DII 9 (1992) ほか多数。

Tomoharu Nagao et al.: Automatic Construction of Image Transformation Processes Using Feature Selection Network, First KOREA-JAPAN Joint Conf. on Computer Vision (1991) ほか多数。

長尾智晴 他 : 「ジェネティックアルゴリズム (遺伝的アルゴリズム)」 昭晃堂, 1993 ほか多数。

30年以上に渡る研究実績

XAI ユーザのニーズに合わせた開発



	項目	内容	成果
1	AI基盤技術	XAIの学術研究	理論, 手法, 学会発表, 論文, 特許.
2	産業応用基盤	ユーザー向けのソフト開発	一般向けソフト, データベースなど.
3	産業応用	miRNA等ヘルスケア応用	健康推進・がんリスク/未病対策事業
		製造業・金融業他	関連企業での実施, YNU発AIベンチャー活動

本研究の特許出願・学会発表など

: 本事業の先導研究での成果 (2019年度)
 : 本事業の1年目での成果 (2020年度)
 : 本事業の2年目での成果 (2021年度)

● 特許出願 (出願中9件, 取得済み10件)

No.	出願者	出願番号	種別	出願日	状態	名称	内容
特許0	横浜国立大学	PCT/JP2018/028633	PCT	2018.7.31	出願中	浸透学習法 (特許4の原型)	学習時のみ利用可能な情報を学習できる世界初の深層回路
特許1	横浜国立大学	特願2019-137811	国内	2019.7.26	出願中	DDN2EME	深層回路を線形回路に変換してグラフやルール, 文章で説明
特許2	横浜国立大学	特願2020-75691	国内	2020.4.21	出願中	miRNA可視化・影響度表示	miRNAのがんリスクの可視化・各がんへの影響度表示
特許3	横浜国立大学	特願2020-75692	国内	2020.4.21	出願中	状態空間による予測	複雑系を状態空間中の点で可視化, 軌跡で将来変動予測
特許4	横浜国立大学	特願2021-032097	国内	2021.3.1	出願中	精度随時向上法	潜在変数を追加しても構造は変化させない深層回路
特許5	横浜国立大学	特願2021-041855	国内	2021.3.15	出願中	非線形パラメータ共有NN	AI組込のために活性化関数をB-splineで実現する手法
特許6	キューピー(株)	PCT/JP2021/016184	PCT	2021.4.21	出願中	miRNA画像化	miRNAを画像に変換してCNNの手法を適用可にする方式
特許7	横浜国立大学	特願2021-182822	国内	2021.11.9	出願中	時間均等状態空間法	状態空間中の状態間の距離が推移時間に比例させる解析法
特許8	横浜国立大学	特願2021-1197852	国内	2021.12.6	出願中	DNN2LC	小型機器へのAI導入のためのDNNの論理回路への変換法

● 国際会議・論文・プレスリリースなど

発表者	所属	タイトル	発表誌名	発表年月
キューピー株式会社		キューピーが、がん発症リスク判定サービス事業化を目指すワケ	Beyond Health (日経BP)	2020.9
大塚蔵高、落谷孝広	QP・TMU	Possible connection between diet and microRNA in cancer scenario	Seminars in Cancer Biology	2021. 1
キューピー株式会社		論文『がんにおける食とマイクロRNAの関係性とその可能性』が国際学術誌 Seminars in Cancer Biologyに掲載	プレスリリース	2021.2
キューピー株式会社		Preventing disease before it starts Natureの会誌・Webの両方に本取組みが掲載	Nature誌	2021.7
荒井 敏、白川真一、長尾智晴	YNU	Non-strict Attentional Region Annotation to Improve Image Classification Accuracy	IEEE SMC-2021	2021. 10
小林雅幸、白川真一、長尾智晴	YNU	Auxiliary Data Selection in Percolative Learning Method for Improving Neural Network Performance	ICAART 2022	2022. 2



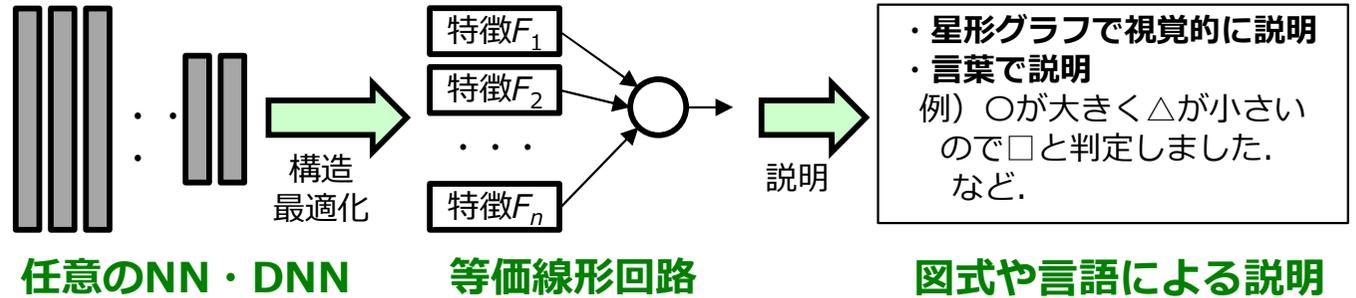
● 長尾によるXAI関連招待講演・解説など

➤ 情報処理学会誌(2022年8月号) XAI特集での全体解説など, 多数.

成果例1: 深層回路(DNN)の線形回路化による説明

- **特許** : 特願2019-137811 (横浜国立大学)
- **内容** : 深層回路を**線形回路に変換**して図や言葉で説明することができる手法。
- **方法** :

1. 進化計算・勾配降下などを用いて**入力・中間ユニットの削減**による**構造の最適化**を行う。
2. 特徴を残った**入力データの数式**として**進化計算法**により数式化。



● 応用例 :

➢ 血液中に約2,500種あるマイクロRNAの値からがんのリスクを判定する

元のNNの構造

非線形 (解釈が難しい) → 線形 (出力層の手前の値をGP(数式)で近似)

$$f_1 = \tanh(-mR_{1443})$$

$$f_2 = \tanh(mR_{1443} + 3 \times mR_{159})$$

$$f_3 = \tanh(\min(mR_{159}, \min(mR_{159}, mR_{1443}) + mR_{51}) + mR_{51})$$

各特徴量を進化計算によって入力変数の単純な式で近似して理解し易くする

適正值 ローリスク (blue area)
 異常値 ハイリスク (red area)

$F_1 \sim F_5$: 特徴量
 $A \sim K$: マイクロRNA

特徴量の線形和で直感的に理解して頂く。

判定結果 マイクロRNAスコア52点(あなたの年齢の平均: 64点)

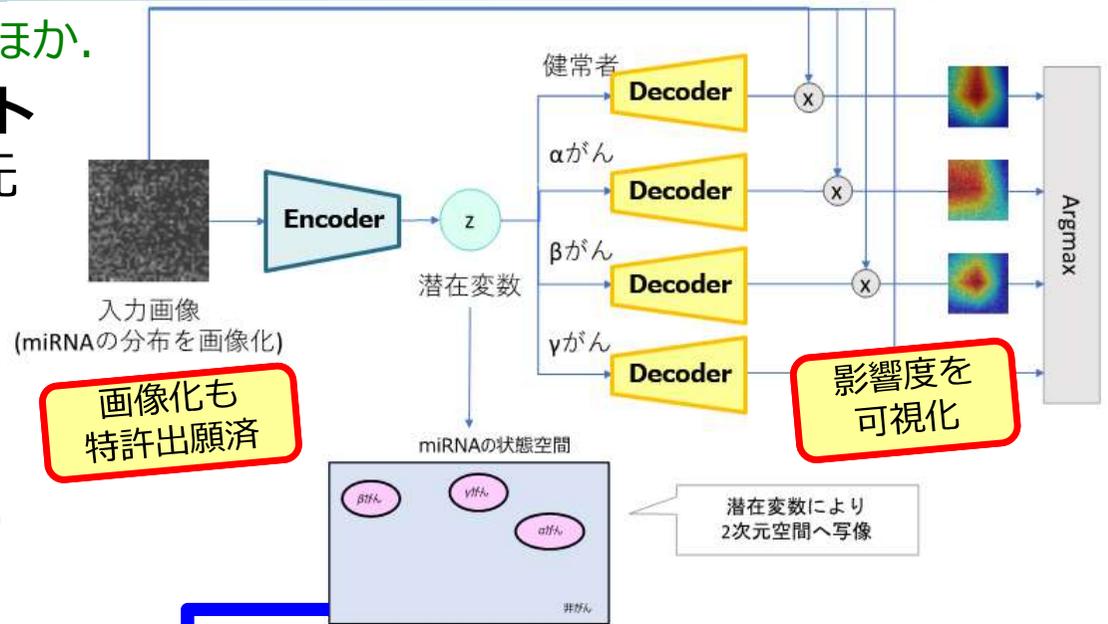
- がんへの抵抗力を示すマイクロRNAスコアが低くなっております。
- 特に前立腺がんとの関係が深い F_2 の数値が低い状態です。
- F_2 はマイクロRNA B~Eの数値が影響しています。

精度劣化なく1/100以下(0.8%)の構造圧縮を既に達成!

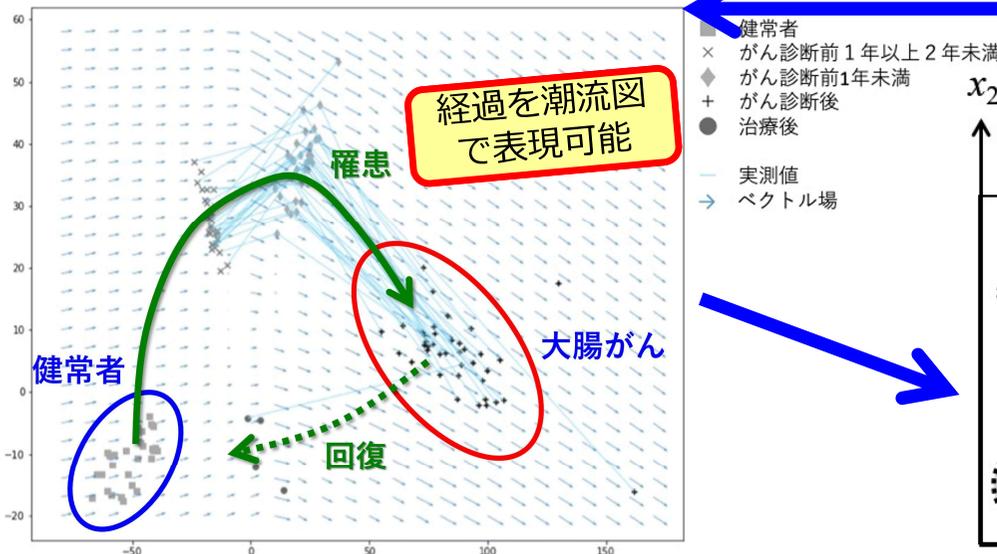
	元のNN	単純化後
入力素子数	2,565	35
中間素子数	3,072	11
誤差率	6.34%	4.20%

成果例2: 状態空間による系の記述と変動予測

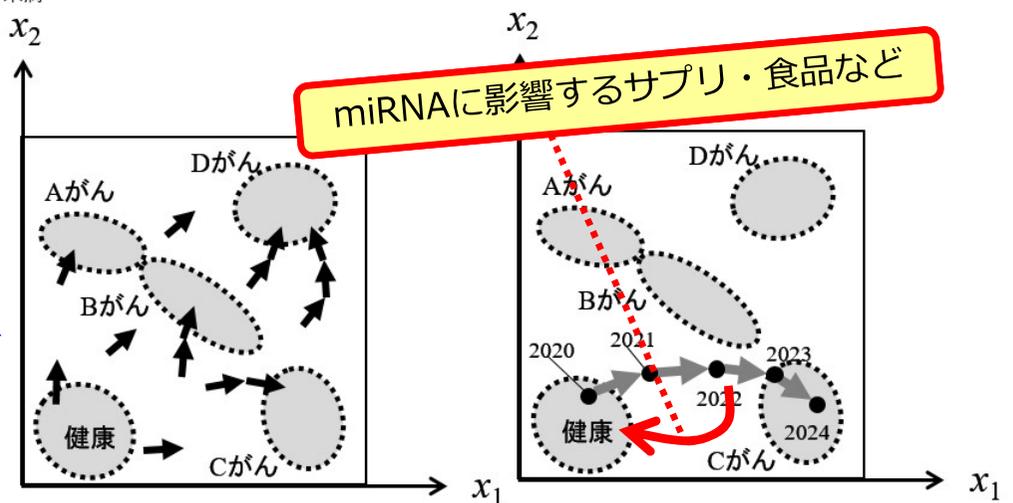
- **特許**: 特願2020-75691 (横浜国立大学)ほか.
- **内容**: miRNAのがんへの寄与度をヒートマップ表示するとともに, 2次元の状態空間で将来変動を予測.
- **応用例**:
 - 血液中に約2,500種あるマイクロRNA (miRNA)分布から健康状態を推定し, **将来がんにならなくて済む健康社会の実現を目指す** (市場規模: 数千億円).
 - プラント・機械の異常の予測も可能.



共同研究機関のキューピー(株)が2025年の事業化を目指している



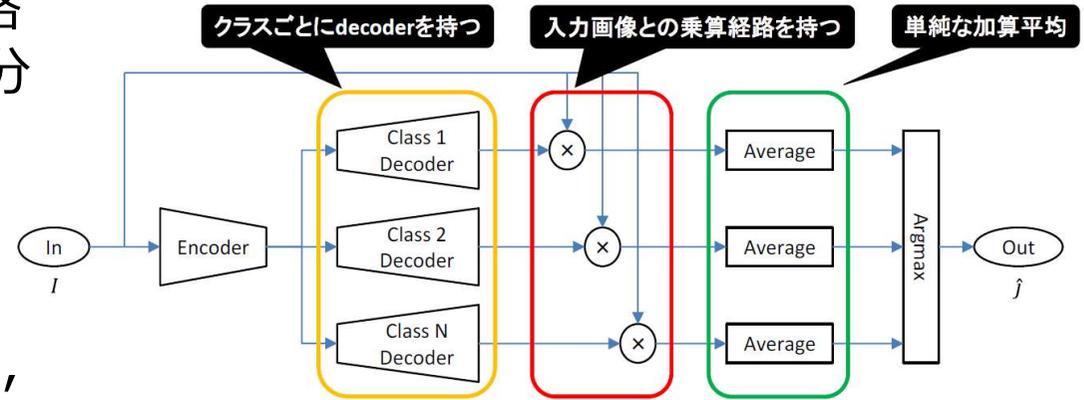
実測値のプロット例



将来の未病対策のイメージ

成果例3: 部品の欠陥検出の判断根拠領域の可視化

- **内容** : 内部に可視化部分を含む深層回路であるGCMによって、画像のクラス分類の際の根拠領域の検出を行なう。
- **応用例** :
 - 協力企業における**画像による欠陥検出**処理に適用して有効性を検証した結果、**Grad-CAM などの従来法より正確に欠陥領域を検出できる**ことを確認した。



GCM: Generative Contribution Mappings

<p>GCM (提案手法) による可視化</p>	<p>粒状欠陥 線状欠陥 POC先企業様の画像につき公開不可</p>	<p>検出された判断根拠領域</p>	<p>全ての欠陥を過不足なく検出</p>	<p>線状の欠陥も良好に検出!</p>
	<p>同じ画像 画像 1</p>	<p>同じ画像 画像 2</p>		
<p>Grad-CAM (従来法) による可視化</p>	<p>無関係な反応 正解 無関係な反応 POC先企業様の画像につき公開不可</p>	<p>線状に検出できていない</p>	<p>線状欠陥が分離</p>	<p>無関係な反応</p>
	<p>同じ画像 画像 1</p>	<p>同じ画像 画像 2</p>		

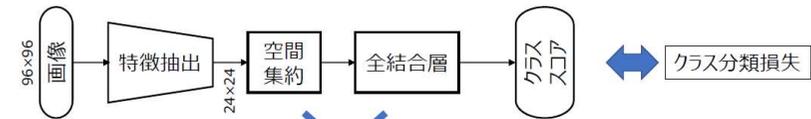
成果例4: 人の判断根拠を利用したCNN精度向上

CNN: Convolutional Neural Network

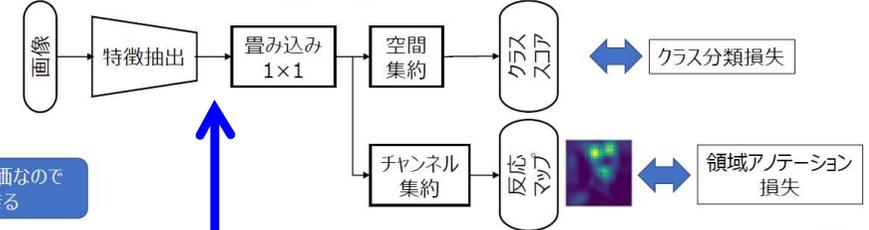
- **内容** : MTurkで収集した人の画像分類の判断領域を用いてCNNの精度を改善.



一般的なクラス分類CNNモデル



今回のモデル



クラス分類部分は等価なので相互に変換できる

- **方法** : 右図の構造の深層回路で実現.

- **実験例** :

- アノテーションの利用方法を変えて試した.
- いずれもベースラインより認識精度が向上した.

欠陥検査など産業用画像処理などや医用画像処理でも有効と考えられる.

ベースラインCNN (WRN28-2)	アノテーションの種類	アノテーション1回分だけ使用	アノテーション3回の平均を使用	領域アノテーションを正規分布近似	円形領域でサイズ固定
83.40%	精度	85.76%	86.68%	86.50%	86.95%
	精度向上幅	+2.36%	+3.28%	+3.10%	+3.55%
 入力画像 従来法	領域アノテーション (目標)				
	CNNの反応マップ (実際)				

人の知識で精度を向上!

成果例5: 入力変数の選択による深層回路の最適化

- **内容** : 論文から得られた知見(事前情報・経験則など)を入力情報の選択に利用する。
- **方法** : 関連する論文情報がある特徴量, 低コストで取得できるセンサ値などに関する事前情報(人が深層回路の入力として選ばれてほしいと考える特徴量)を参照し, **個々の特徴量の選択され易さを調整する係数を導入した。**

- **実験例** : あるデータの予測値を比較した。

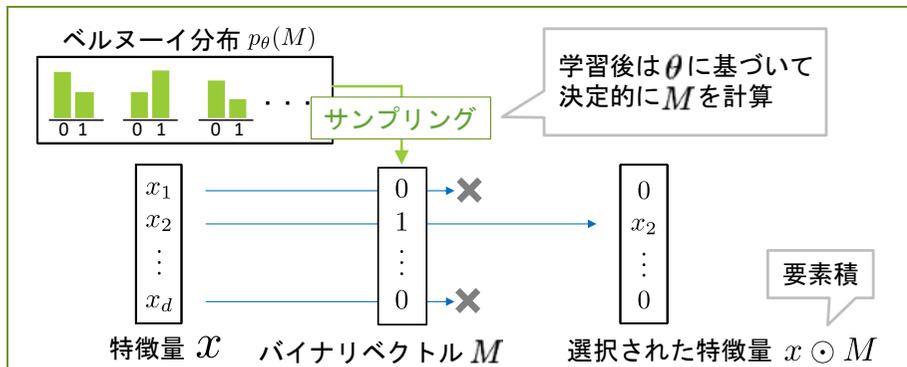
- 人の知識を利用した方が**常に予測精度が上昇するわけでもない点が興味深い。**
- downとなったケースでは, **論文で指摘された特徴量より有効な特徴量が存在する可能性を機械学習が示唆している**と考えることもできる。
- いずれにせよ, これによって**機械学習と人との間の意見交換, 共進化が生じることが期待される。**

人と機械の対話が生じる

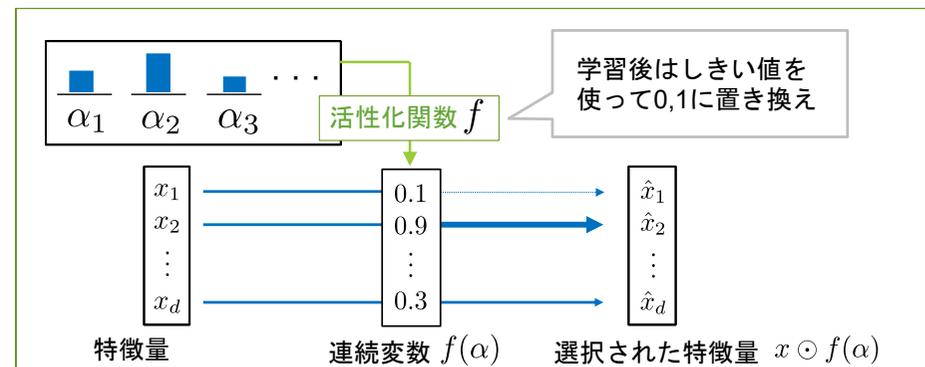
	事前情報	総選択特徴数	重要な特徴の選択率 (%)	予測精度 (%)	
連続緩和 シグモイド	なし	495	22.6	95.3	↓ down
	あり	407	50.1	94.4	
連続緩和 温度付きシグモイド	なし	448	21.7	95.4	↕ even
	あり	454	50.2	95.4	
連続緩和 ガンベルシグモイド	なし	551	24.5	94.8	↓ up
	あり	474	47.7	95.3	
確率緩和	なし	451	22.2	96.1	↓ down
	あり	176	40.9	95.2	

人の知識導入による変化

確率緩和 : 離散変数の確率分布の分布パラメータを最適化



連続緩和 : 離散変数を連続変数に置き換え最適化



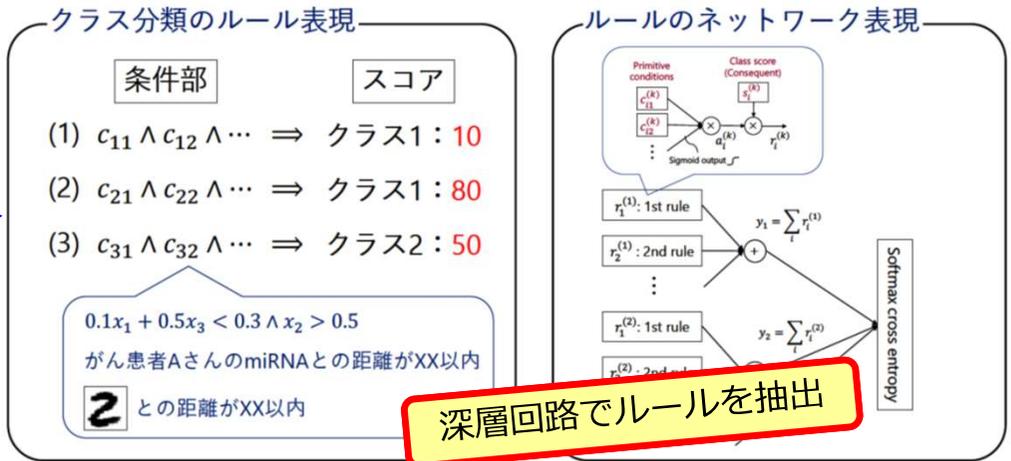
成果例6: 勾配ベースでのルールの獲得

- **内容**: クラス分類のルールはネットワークでも表現することもできることに着目し, 深層回路でルールを抽出する手法を開発中である。

- **実験例**:

- **MNIST**では各クラスの数少ないパターン(テンプレート)と比較するだけで, 十分認識できることがわかった(下図左).

- 一方, 一般物体認識系のデータ(**CIFAR10**)ではまだ難しく, 現在研究中である(下図右).



評価実験

MNIST:
ルール表示

入力:

クラススコア: 29.9757
予測クラス: 0

テンプレート	しきい値	クラススコア	テンプレート	しきい値	クラススコア
(1) ZNCC:	> 0.7628	⇒ 20.4907	(11) ZNCC:	> 0.7347	⇒ -4.3910
(2) ZNCC:	> 0.7345	⇒ -4.8849	(12) ZNCC:	> 0.7347	⇒ -6.1374
(3) ZNCC:	> 0.8009	⇒ -4.8849	(13) ZNCC:	> 0.7347	⇒ -6.1374
(4) ZNCC:	> 0.7396	⇒ 17.4510	(14) ZNCC:	> 0.7350	⇒ -7.0704
(5) ZNCC:	> 0.7345	⇒ 12.5247	(15) ZNCC:	> 0.7347	⇒ -9.5222
(6) ZNCC:	> 0.7345	⇒ 8.2988	(16) ZNCC:	> 0.7414	⇒ -14.2085
(7) ZNCC:	> 0.7345	⇒ 7.9458	(17) ZNCC:	> 0.7355	⇒ -14.3589
(8) ZNCC:	> 0.7345	⇒ 5.3562	(18) ZNCC:	> 0.8612	⇒ -15.1814
(9) ZNCC:	> 0.7345	⇒ 4.2050	(19) ZNCC:	> 0.8134	⇒ -16.1389
(10) ZNCC:	> 0.7347	⇒ -3.7241	(20) ZNCC:	> 0.7433	⇒ -16.3831

深層回路を単純なパターンマッチングの問題に変換!

飛行機
自動車
鳥
猫
鹿
犬
カエル
馬
船
トラック

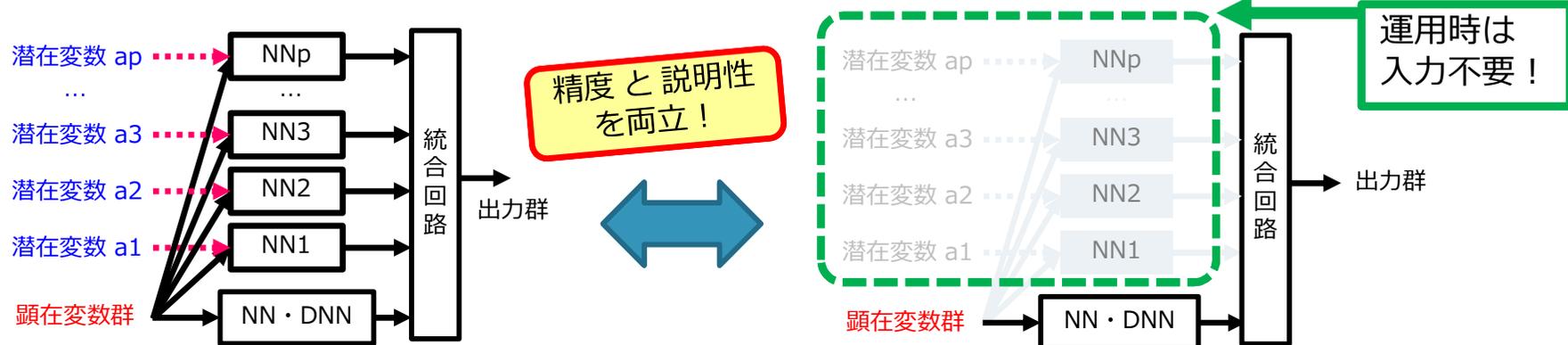
学習されたテンプレート 上位10件

手書き文字データベースMNISTにおけるゼロの認識

一般画像データCIFAR10の場合

成果例7: 潜在変数による精度と説明性の両立

- **特許** : 特願 2021-032097 (横浜国立大学)
- **内容** : 研究代表者が先に開発した**浸透学習法**(PCT/JP2018/028633)を応用して, 説明変数を**顕在変数**と**潜在変数**に分け、**学習時のみ潜在変数を使う深層学習方法**を開発し、既存のDNNに追加的に**潜在変数**を加えて精度を向上させる方式を開発した。



学習時 : 潜在変数も考慮して精度を向上

運用時 : 顕在変数だけで説明性が高い

- **適用例** : 様々な応用が可能!

- **がんリスク・生活習慣病等の判定** (理由: 一般に専用装置・人・時間のコストが高い)
 - **顕在変数** = 計測が容易なバイタルデータ, **潜在変数** = 計測が大変なバイタルデータ
- **時系列信号の将来変動の予測** (理由: 入力変数の十分性が保証できていない)
 - **顕在変数** = 株価・為替など, **潜在変数** = 有料の海外ニュースや高コストの情報。
- **自動運転車両の自動制御** (理由: 製品に搭載可能なセンサ数にはコストの制約が厳しい)
 - **顕在変数** = 市販車両のセンサ, **潜在変数** = 試験車両に付けられるだけ付けたセンサ群。
- **人の表情からの感性計測** (理由: 利用時に使えるセンサには制限がある)
 - **顕在変数** = カメラ映像, **潜在変数** = fMRI・脳波・心拍・血圧などの計測が大変な情報。

成果の産業応用について

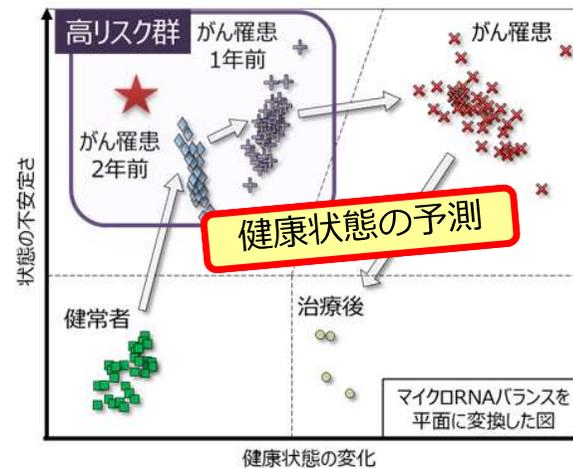
1. ヘルスケア分野

- 血液中のmiRNA分布を用いた健康状態をモニタリングして、現在および将来の**がんリスクを判定し**、がんにならないための食生活改善を指導する。
- キューピー株式会社を中心に社会実装予定。

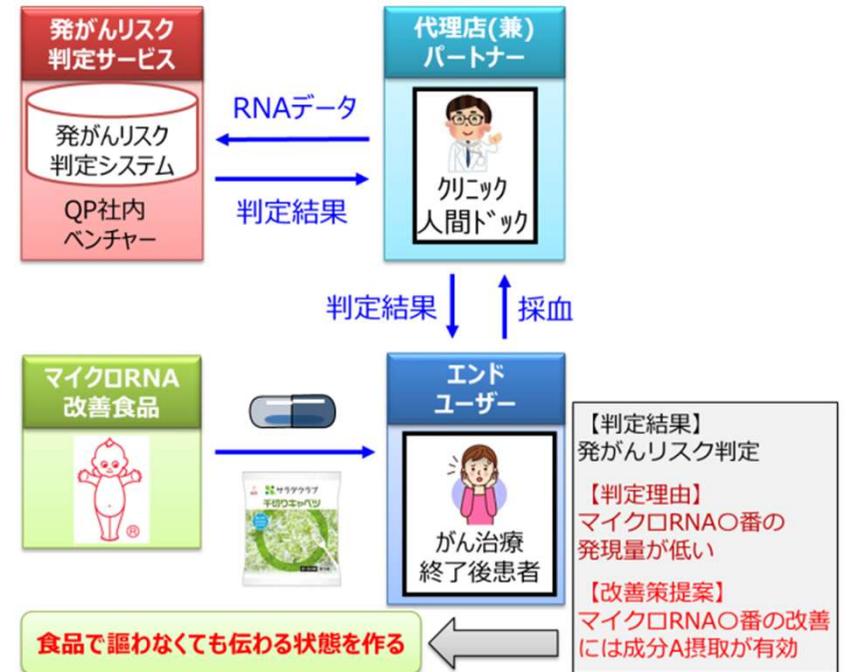
健康状態の把握

種別	将来の発がんリスク				
	ローリスク	←→			ハイリスク
	A	B	C	D	E
がん全般 マイクロRNAランク			★		
大腸がん		★			
胃がん					★
食道がん	★				
乳がん	★				
肺がん		★			

提示方法 (患者用)



提示方法 (医師用)



ビジネスモデル案

2. ヘルスケア分野以外 (製造業・金融業など)

- XAI・CAI 技術のヘルスケア分野以外への応用。
- 当面は**画像と時系列信号の異常検知・予測・根拠説明**などの業務を想定して企業や大学に身近なビジネス展開をする予定。
- 横浜国立大学発の新AIベンチャーを中心に社会実装予定。



発表内容（再出）

- 説明可能AI (**XAI**)・共進化AI (**CAI**)

XAI: eXplainable AI,

CAI: Co-evolutional AI

- 当研究の概要
- 当研究の成果例
- 成果の産業応用について

ご清聴ありがとうございました。

今後の成果にご期待下さい。

皆様の益々のご発展をお祈りします。

