

説明可能なAIによる 「信頼できる診断と予測」の実現

株式会社富士通研究所
デジタル革新コアユニット
プリンシパルエキスパート 渡部 勇

Copyright 2019 FUJITSU LABORATORIES LTD

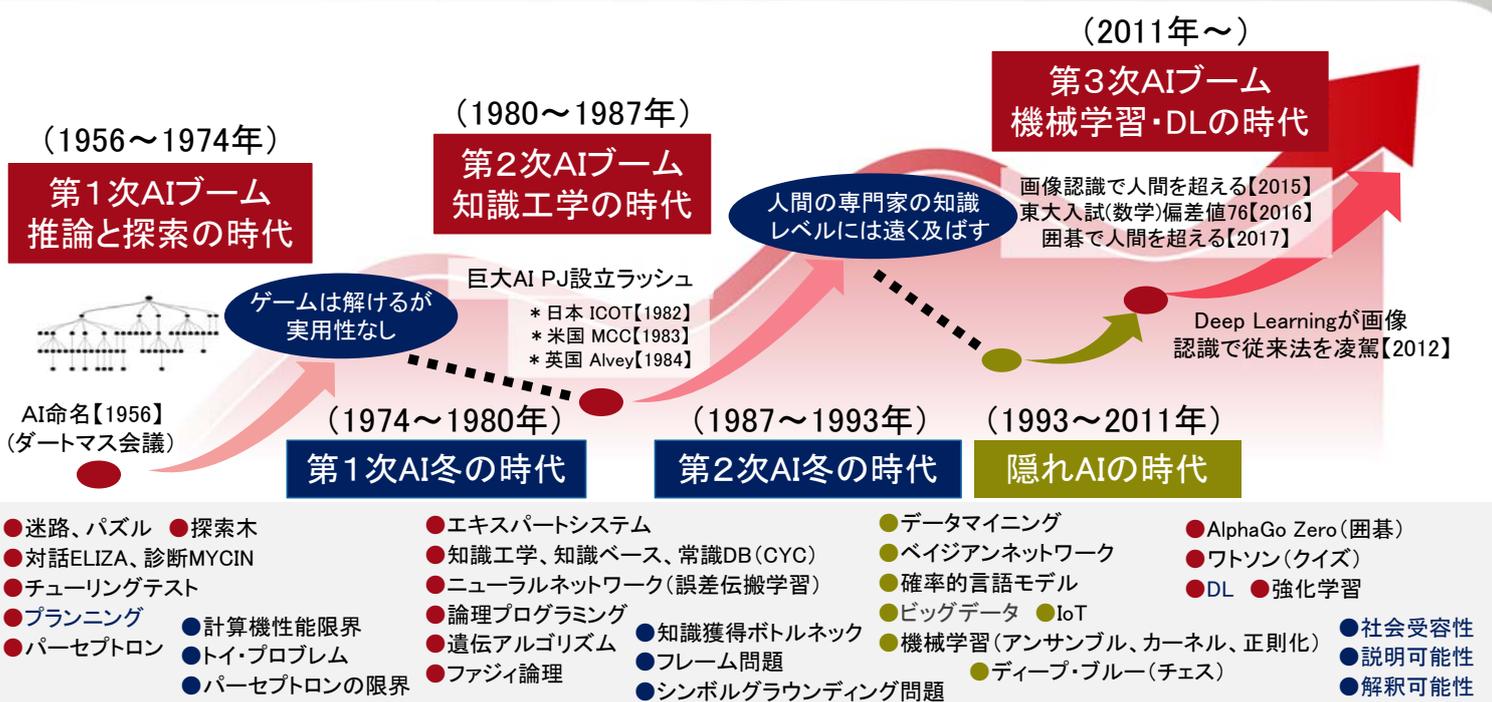
本日の講演内容



- 【第一部】人工知能技術の動向
 - 人工知能の歴史、機械学習、Deep Learning、残された課題
- 【第二部】説明可能なAI
 - 説明可能なAIの必要性、動向、アプローチ、適用事例
- 【第三部】予知保全のためのAI技術(これまでの取り組み)
 - 予知保全のトレンド、目的、技術
 - 異常・故障の原因分析(テキストマイニングによる説明可能なAI)
 - 劣化量推定／予測(近似モデルによる説明可能なAI)
- 【第四部】「信頼できる診断と予測」の実現に向けて
 - 予知保全の課題と解決方法 → 説明可能なAI／解釈可能な機械学習の活用
ハイブリッドアプローチ(知識+DL・機械学習)

【第一部】人工知能技術の動向

人工知能の歴史



機械学習とは？

■ 人工知能における機械学習 (Wikipediaでの定義)

- 人工知能における研究課題の一つで、人間が自然に行っている学習能力と同様の機能をコンピュータで実現しようとする技術・手法
- 明示的にプログラムしなくても学習する能力をコンピュータに与える研究分野



- 人工知能の文脈での定義は上記だ、何でもできそう(かつ抽象的)だが、データ分析における「機械学習」はもっと限定的で具体的

■ 機械学習でできること(ざっくり言うと…)

- 統計的なモデルのパラメータをデータから推定
 - ビールの売り上げ
= CM量 × 係数1 + 気温 × 係数2 + 価格 × 係数3 + …
- 統計的なモデルを用いて下記を実現
 - ビールの売り上げに何がどの程度効いているのか？ → 説明的分析
 - 来週はビールがどれだけ売れるのか？ → 予測的分析
 - 利益を最大化するには、CMと価格をどうすれば良いか？ → 指示的分析
- 機械学習の方式とタスク
 - 教師あり学習: 判別(分類)、回帰(数値予測)
 - 教師なし学習: クラスタリング、次元圧縮、異常検知

Deep Learningとは？

- 機械学習の一種: 大量のデータから学習
- 超多層(=ディープ)のニューラルネットワーク
- 特徴 = 超多層にすることで特徴を自動抽出(表現学習)



写真を見て「犬」と分かるには？

■ (例) 犬を認識する画像認識処理

■ 古典的な機械学習

- 画像から犬を判別する「特徴」を人が教え込み、コンピュータが判断

① 4本足



② 鼻が尖っている



③ 毛で覆われている



犬には他にどのような特徴がある？
キツネとの違いがわかる特徴は？



犬の特徴をすべて言葉にする
(数量化する)のは難しい

人が特徴量を教え込むことの限界

■ 人が特徴量を考え出し、教え込むことは**すごく大変** → 特徴量工学

- 漏れが生じやすく、認識率が低くなってしまう

① 4本足？



② 鼻が尖っている？



③ 毛で覆われている？



人がすべての犬種、動き、状態などを
加味した特徴量を教えることは質・量とも難しい

Deep Learningの特徴＝特徴量を見つける

■ 人が機械に教えなくても、大量画像から特徴量を自動抽出



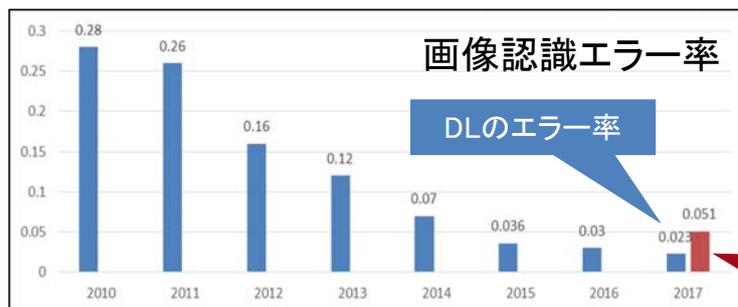
Deep Learningによる画像認識

■ ヒトの脳を模したニューラルネットワークの最新技術

- 特徴点を自動抽出し、様々なパターン認識のコンペティションで桁違いの精度。画像認識、音声認識、など

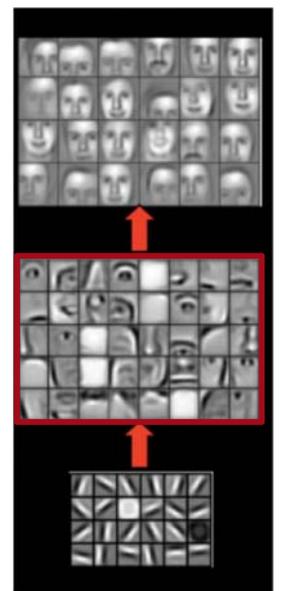
■ ILSVRC (ImageNet Large Scale Visual Recognition Challenge)

- 2012年、トロント大が Deep Learning を利用し、圧勝
- 以後、全チームがDLを利用し、2017年、2.3%のエラー率



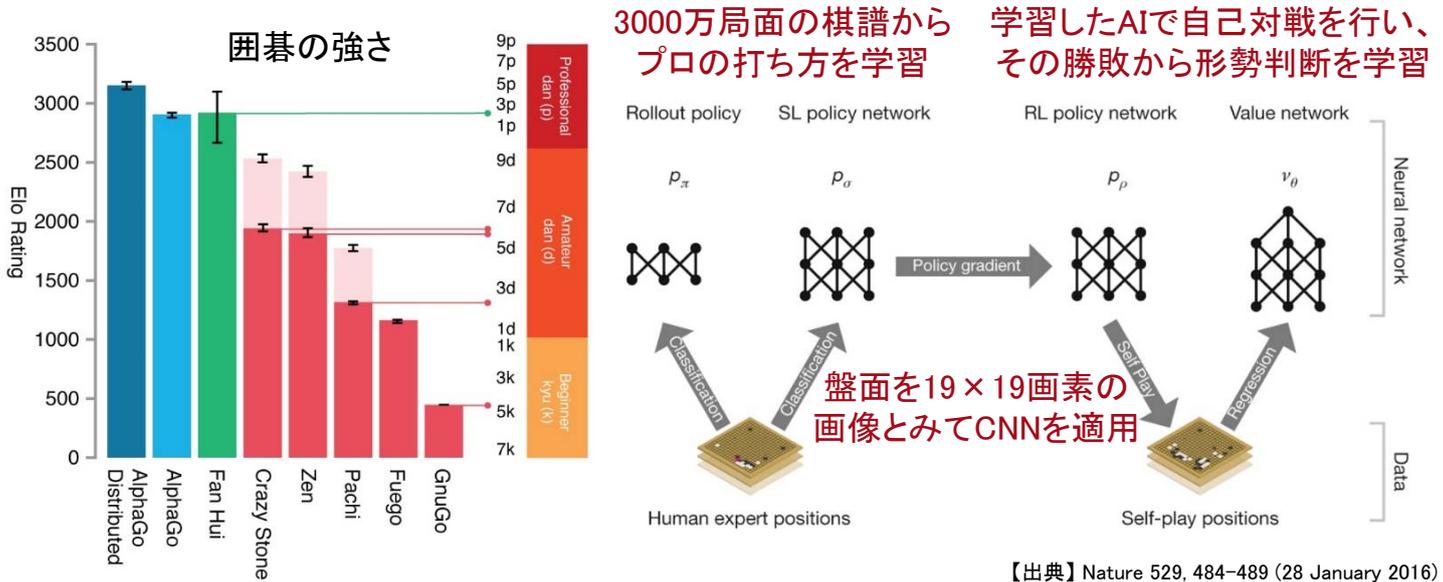
教師を超えられる? → No
人間のエラー率は0% → No

専門知識を持った
(訓練された)
人間のエラー率



顔を認識に必要な特徴を自動的に学習している

■ 人手での設計が難しかった形勢判断をDeep Learning+強化学習で実現



第一部のまとめ①: 課題は解決できた?

■ AI冬の時代に至った課題の解決度

AIの課題	現状での解決度
計算機性能限界	◎ 高性能化(CPU/メモリ)、並列化、高速化アルゴリズム
トイ・プロブレム	◎ DLIによる画像認識・音声認識(実用レベルの精度)
	△ 深層強化学習(囲碁、TVゲームなどの限られた世界)
パーセプトロンの限界	◎ DL・機械学習(複雑で非線形なモデル)
知識獲得ボトルネック	△ 集合知、ナレッジグラフ
フレーム問題、シンボルグラウンディング問題	× (DLにより、部分的に解決済み/近々解決可能という意見あり)

■ できるようになったこと

- パターン認識、制御
- 閉じた世界での問題解決



■ まだ十分にできていないこと

- 意味理解、知識処理、課題発見、問題解決、...
- 実世界での問題解決(現状はタスク依存で限定的)

第一部のまとめ②: 第3次AI冬の時代は来る? FUJITSU

■ 何でもできる「強いAI」「汎用型AI」の実現はだいぶ先

■ 「ロボットは東大に入れるか」プロジェクト

- 2021年度の東大二次試験突破を目指し、NIIの主導で2011年にスタート
- 2016年にセンター試験模試で好成績を残したが、「AIは意味を理解できない」限界のため凍結

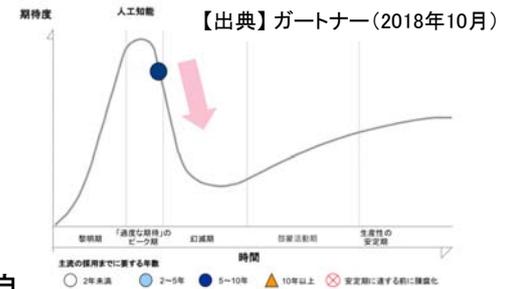
■ 特定のタスクに特化した「弱いAI」「特化型AI」の実現と実用化

■ ガートナーのハイプサイクル(技術の発展サイクル)

- AIは「過度な期待」のピーク期から幻滅期へ移行
- 研究段階から実用段階へシフト(夢から現実へ)

■ 研究のフォーカス

- どのようなタスクに適用可能なのか? → **予知保全**
- 実世界の問題を解くためには何が足りないのか? → **技術改良**
- 社会実装を進めるためには何が足りないのか? → **説明可能なAI**



【第二部】説明可能なAI

説明可能なAIの必要性

- Deep LearningはAIのブレイクスルー技術
しかし、ブラックボックスのため社会実装できる領域が極めて限定的

人間が



信頼できる



理解できる



管理できる



説明可能なAI
(XAI: eXplainable AI)

1

説明責任を果たせる

2

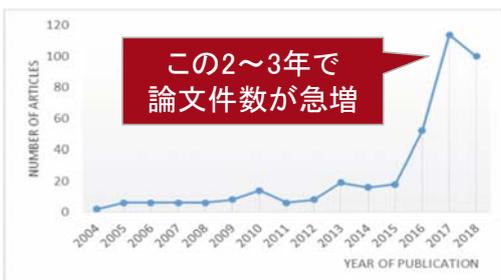
新しい発見ができる

3

AIを改善できる

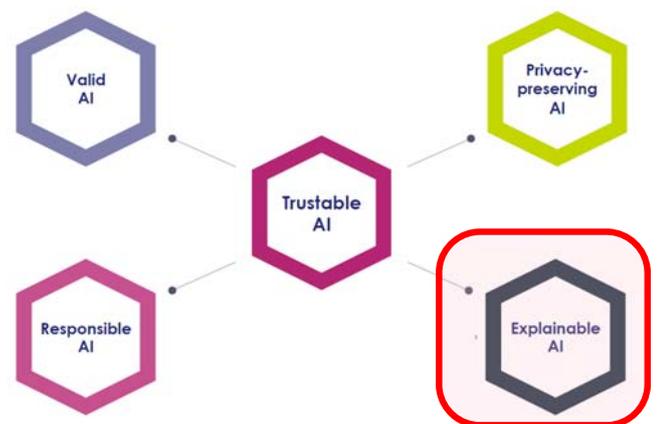
説明可能なAIの動向

- 社会的な要請・背景
 - 【米国】説明可能なAI(DARPA、2016年)
 - 【EU】一般データ保護規則(GDPR、2016年)
 - 【日本】AI利活用原則案(総務省、2018年)
- 学会での位置づけ
 - 著名な国際会議の主要トピックス



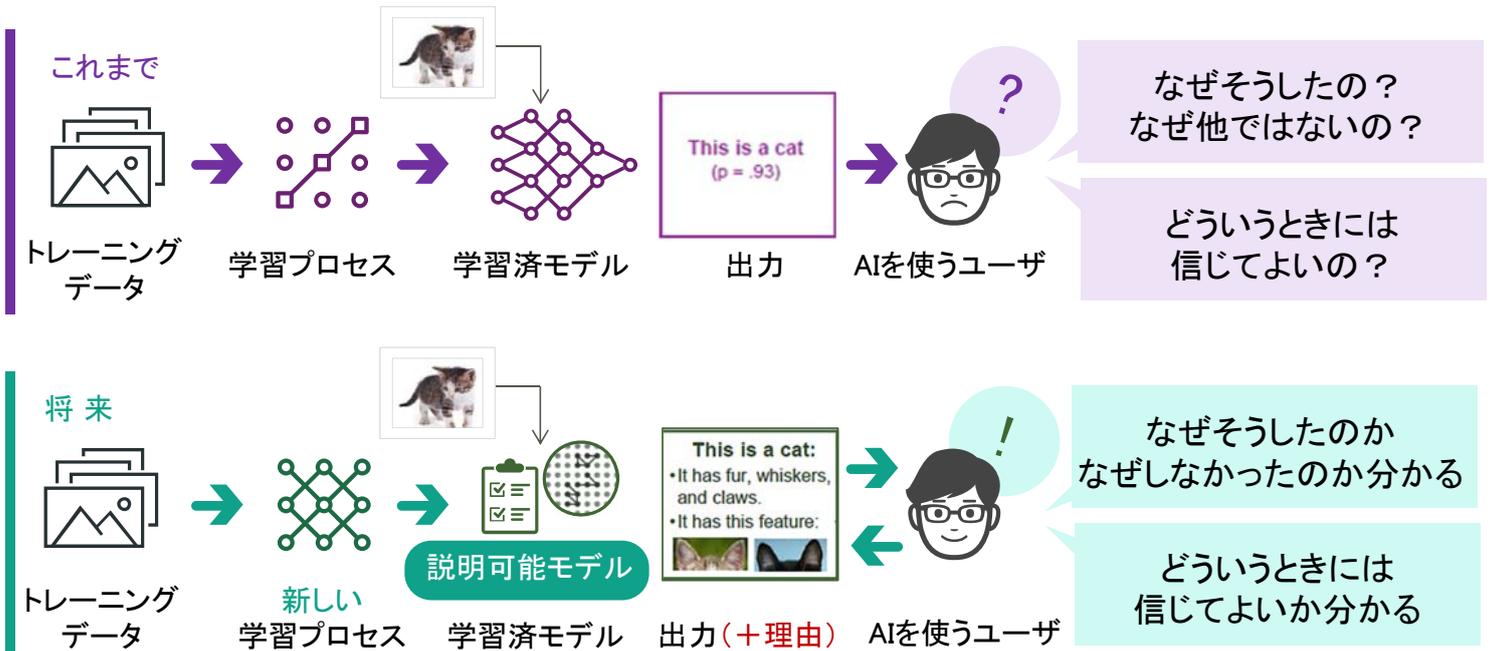
【出典】 Peeking inside the black-box: A survey on Explainable Artificial Intelligence(2018)

- 産業界でのAIの採用要件
 - 「信頼できるAI」の基本構成要素



【出典】 On the Role of Trust and Explanation for AI Adoption in Industry(2019)

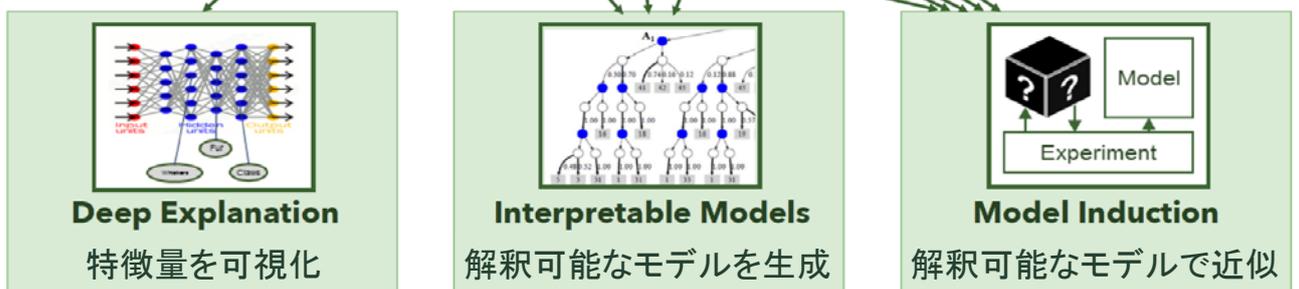
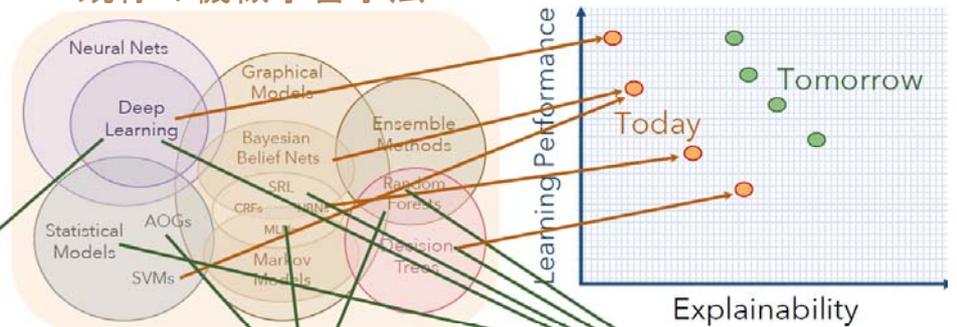
説明可能なAIの目的



説明可能なAIのアプローチ (DARPAの定義)

既存の機械学習手法

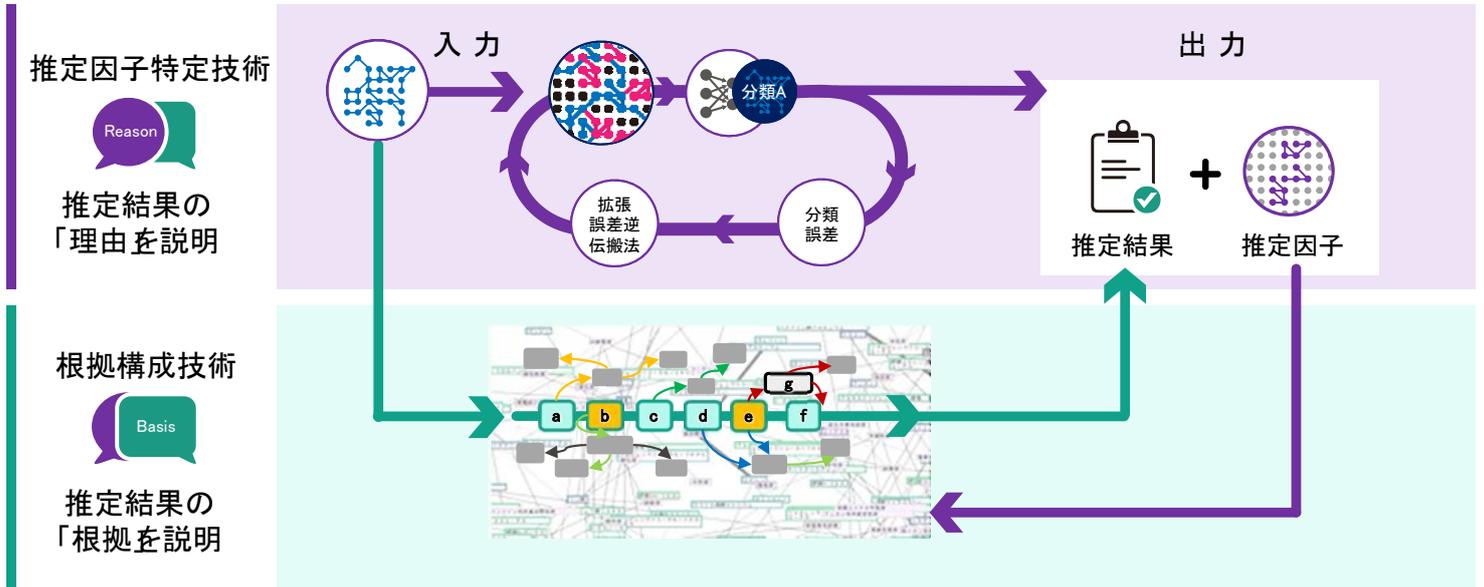
【説明可能なAIの目的】
高性能(精度)と高説明性を
両立する機械学習技術の確立



富士通の提案する「説明可能なAI」



■ 2つの説明を行う全く新しいAI: 結果の「理由」と「根拠」を説明



ゲノム医療 (Genomic Medicine)



富士通の「説明可能なAI」でゲノム医療における分析・診断・治療判断の時間を大幅に短縮



健康経営

富士通社内部門

過去の出勤簿から3か月後の健康変化を予測し理由を説明。早期対応で休職者を削減する

2018下期より、予測機能を先行してサービス開始予定



技術

グラフ構造化

DeepTensor

予測 3か月以内の
体調不良

専門家が適切な対策を立てられるように支援



効果

データの分析作業が1/10に

- 84%の予測精度を実現(既存手法:66%)
- 外した16%: 休職予備軍の可能性

早期に、発症時期と対策を伴う予測が可能に

- 専門家も見つけづらいケースを早期に発見
休職回避の確率・施策の有効性を高める



投融资業務支援

取引履歴やメール・通話履歴へDeep Tensor®適用で 決算書に表れない人や企業の実態を推定 信用度を決算書以外の情報から推定

課題

信用度判定の難しさ

決算書を取得しない
融資先の信用度評価

決算書取得と実態との
時間的ずれによる
信用度の誤評価



効果

- 企業間関係や組織情報が記録されたナレッジグラフを用いて、人間に**解釈可能な言葉**で推定因子を説明
- 複雑なナレッジグラフに**隠れた事実**を推定因子を手がかりとして発見



第二部のまとめ：説明可能なAI＝実用化の鍵

■ 最近の動向：後付けの説明可能性 vs 最初から解釈可能なモデル

【出典】Rudin, Please Stop Explaining Black Box Models for High-Stakes Decision (2018)

【出典】Rudin, Do Simpler Models Exist and How Can We Find Them? (2019)

■ 精度と説明性がトレードオフになるとは限らない(Black Boxと同等精度のWhite Boxが存在)

■ 近似モデル(後付け説明)は不正確な説明を提供する可能性がある

→ 可能な場合は「**解釈可能なモデル**」を採用すべき

■ 目的に応じた説明可能なAIの選定が重要

■ 精度が高ければよい／失うものがない → **Black Boxモデル**

・ マーケティングにおける広告配信の最適化、製造工程における抜き取り検査の最適化

■ 専門家のための判断材料／仮説を提供 → **Grey Boxモデル(後付けの説明可能性)**

・ 医療におけるスクリーニング／診断支援

■ 専門家の肩代わり(非専門家が使う) → **White Boxモデル(最初から解釈可能なモデル)**

【第三部】 予知保全のためのAI技術 (これまでの取り組み)

予知保全の目的とトレンド

■ 予知保全の目的

- 稼働中の設備・機械の状態を診断し、**保全すべき時期を予測**

→ 故障・異常による**ダウンタイムを短縮し、損失と保全コストを削減**

【出典】次世代メンテナンスのあり方 (Roland Berger 2017)

■ 保全方式のトレンド

事後保全

↓
予防保全(時間基準)

↓
予知保全(点検ベース)

↓
予知保全(モニタリング)

	第一世代 1940 1950	第二世代 1960 1970	第三世代 1980 1990 2000-
目的	壊れたものの修理	アベイラビリティ向上 より低コストで設備寿命延長	アベイラビリティ・信頼性・安全性・機器性能向上 ライフサイクルコスト(LCC) 効率の追求 設備寿命延長
保全方式(考え方)	Reactive Maintenance 事後保全(≒SDM) > 故障後に保全	Preventative Maintenance 予防保全 > 故障の発生前に保全	Predictive Maintenance 予知保全 > 予兆を掴んで故障発生を予知し、最適な時期に保全 Proactive Maintenance 積極保全・改良保全 > 故障の根本原因を監視・発見し、事前に除去 Reliability-centered Maintenance (RCM) 信頼性中心保全 > 要求信頼性ごとに保全手法を組み合わせ、LCCを最適化
方法・手段	手作業・人力	Time-based Maintenance 時間基準保全(TBM) > 故障の有無に関係ない定期検査・修繕 > (経験・カンに基づく診断によるCBMも含む)	Condition-based Maintenance 状態基準保全(CBM) > センサネットワーク・状態監視システムとデータ分析に基づく故障予兆診断 > 故障モードと影響の解析・リスク管理分析・アセットマネジメント > 高度ITシステムによるビッグデータ処理 など

予知保全(Predictive Maintenance)

= 状態基準保全(Condition-based Maintenance)

= PHM(Prognostics and Health Management)

予知保全の機能アーキテクチャ

■ 予知保全(PHM)の基本機能

■ Diagnostics(診断)

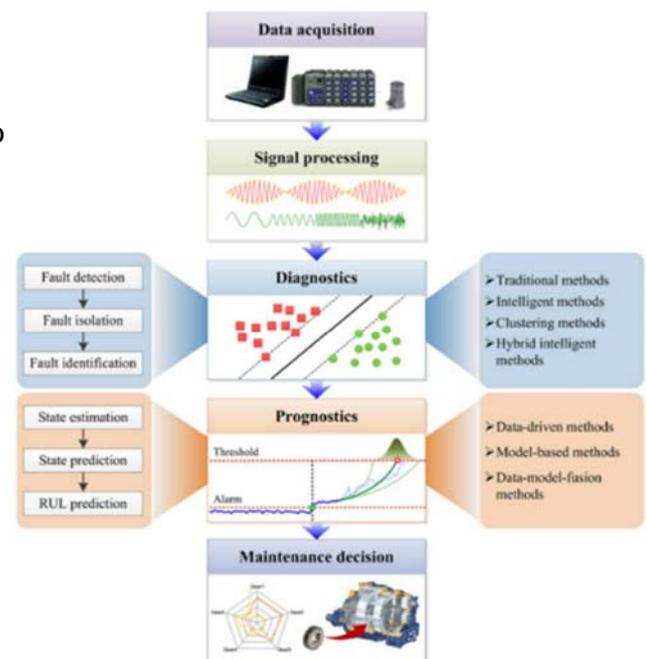
- 機器の状態を把握することで、**異常の予兆**を捉える (Anomaly detection)
- **不具合の種類・箇所**を特定する (Fault Isolation)
- **不具合の程度**を知る (Fault identification)
- **異常の原因**を推定する

■ Prognostics(予測)

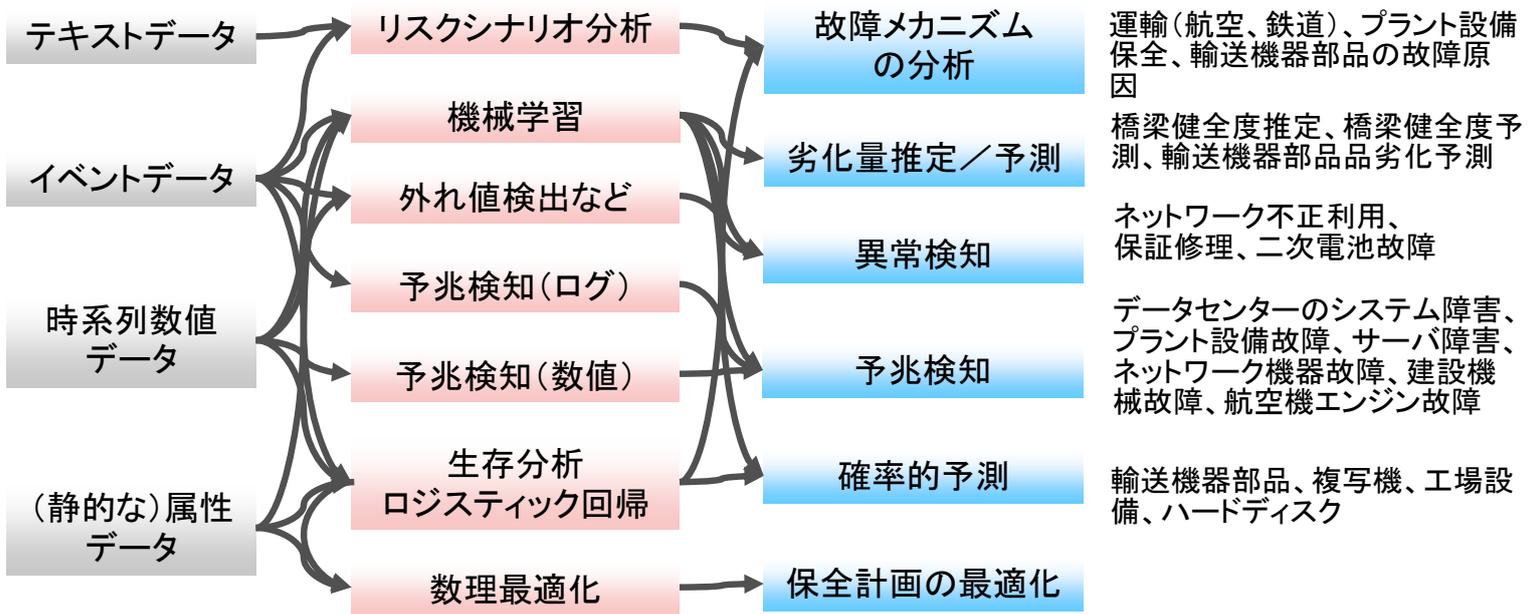
- 推定原因に基づいて劣化進行をモデリングして、将来の状態あるいは**余寿命(RUL: Remaining Useful Life)**を予測する

■ Health Management

- Diagnostics/Prognosticsの結果に基づいて、メンテナンスに関する**適切な意思決定**をする



予知保全のためのAI技術とこれまでの取り組み **FUJITSU**



異常・故障の原因分析 (テキストマイニングによる説明可能なAI)

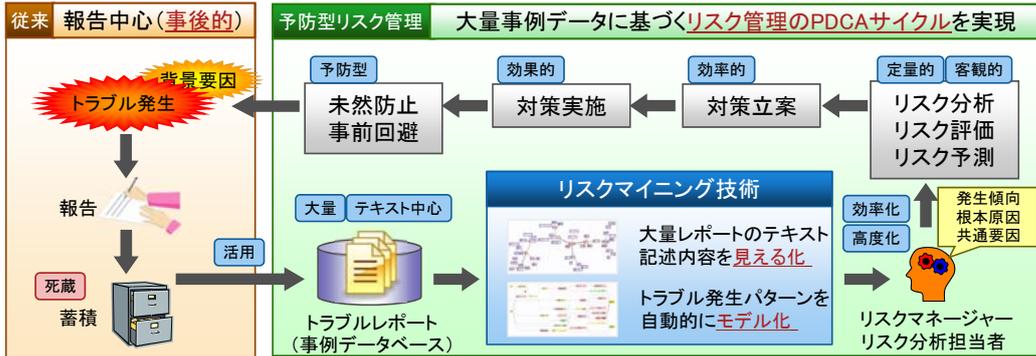
リスクマイニング技術による原因分析

品質・安全性のトラブル → 業界・業種を問わず頻発

- 事業継続や企業経営に多大な影響（航空、鉄道、電力、製造、食品…）
- 過去事例から予見・予防可能（現場では、類似のヒヤリハットが発生）

リスクマイニング技術で実現するリスク管理

- 大量のトラブルレポート(過去事例)を解析し、傾向・要因を分析
- 効果的な対策立案、重大トラブルの未然防止・事前回避を支援



トラブルレポートの例(交通事故)

定型項目

非定型項目(テキスト記述)

定型項目								詳細内容
天候	曜日	時間帯	事故パターン	事故原因	事故原因	負傷状況	事故程度	
曇り	月	午前	自転車+徒歩	歩行者への配慮不足		肩を脱臼	軽傷	月曜日の早朝、私はアルバイトに行くために曇りの中を歩いていた。途中前方から横一列に並んで歩いてくる集団に遭遇した。向かって右端の一人が自転車を運転していたが、その手には傘を前方に突き出す形で握っていた。私はできるだけ道の端を歩いたが、自転車の傘が肩に当たり、横転した。この事故で私は肩を脱臼した。
曇り	火	午前	自転車+徒歩	よそ見運転	歩行者への配慮不足	手首を捻挫	軽傷	火曜日の早朝、友人と徹夜でカラオケをした帰り、曇りの中私だけ自転車で走行していた。歩道側の手で、前方に傘を突き出す形で持ち、走行していたところ、前方から歩行者が来た。あちらが避けるだろうと、そのまま友人と話をしながら走行していたところ、傘が歩行者の肩に接触し、横転した。この事故により、私は手首を捻挫した。
雨	水	午前	徒歩+徒歩	不注意	視界不良	軽い擦り傷	軽傷	水曜日の早朝、私は仕事に行くついでにゴミを出そうと、徒歩でゴミ捨て場に向かった。雨が降っていたため片手で傘をさし、もう一方の手には重いゴミを持っていた。途中、風が向かい風になり、傘を斜め前に突き出して歩行したため、前方から歩行者と衝突し、転倒した。結果、

定型項目だけでは背景要因が分からない！
↓
有効な対策には結びつかない

重要な情報はテキスト記述の中に含まれている
↓
単純に集計することができない

テキストマイニング技術の適用

トラブルレポート(タイタニック号沈没事故)

【経過】不沈の船として世界中から注目を浴びながら、1912年4月10日にタイタニック号は**処女航海に出発した**。予定より**一ヶ月遅れ**で、**出発時間も一時間遅**れていた。2度氷山の警告を受けたが、**深刻に受け止めなかった**。**減速もせずに航海を続行**。14日の11:40、**氷山を450m手前で発見**。**減速しながら舵を切った**が船腹をなでるように**氷山をかすめた**。その際には衝突に気付かなかった乗客もいた。その後**多数区画の損傷**により**大量の浸水**。0:14に初めて救難無線を発す。56海里離れたカルパチア号が受信し、救助に向かった。0:44信号灯を打ち上げたが、近くにいたカリフォルニア号は救難信号とは受け取らず、救助にも向かわなかった。そして2:20沈没。カルパチア号は4:10頃到着。

【原因】氷山に衝突したことがこの事故の直接な原因であるが、氷山の衝突にいたった原因としては、2度にもわたる警告を無視したこと、出航が一ヶ月遅れたために**流水が...**

リスクシナリオ分析

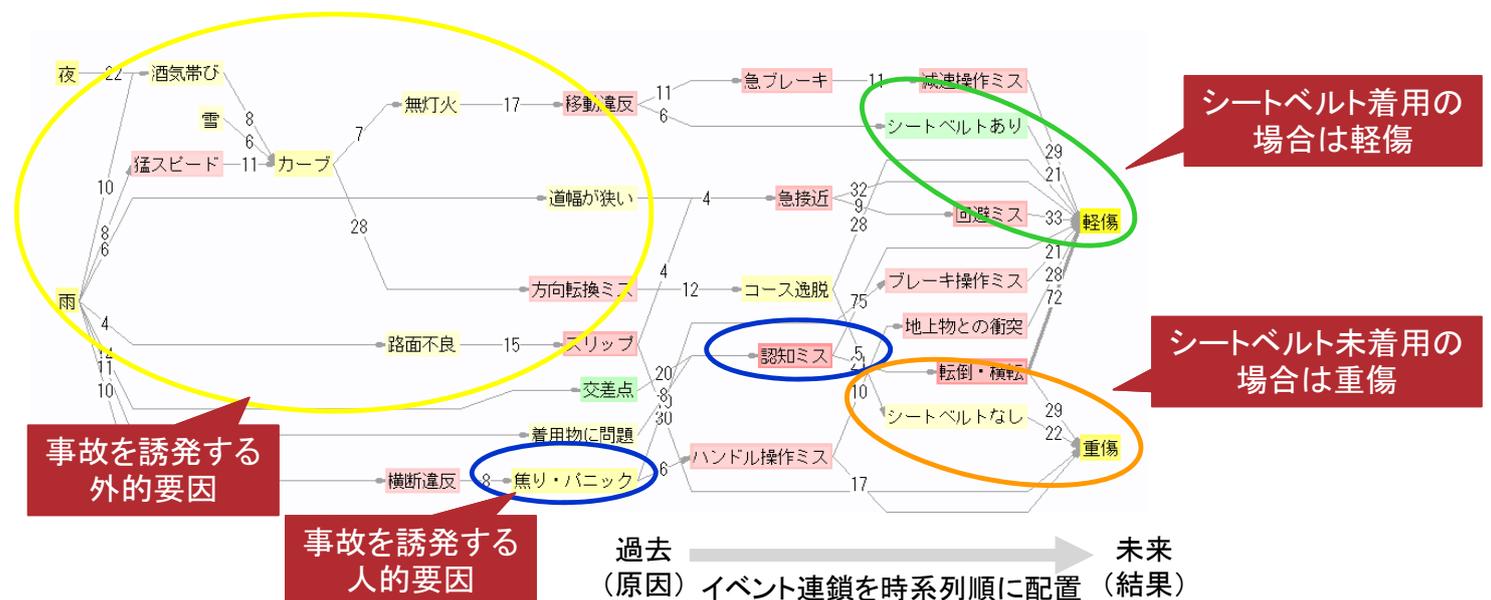
- ステップ1: 文種別の判定
- ↓
- ステップ2: 格フレームの抽出
- ↓
- ステップ3: イベント抽出
- ↓
- ステップ4: 順序関係の抽出

イベント連鎖

- 航海に出発
- ↓
- 遅れが発生
- ↓
- 警告を受信
- ↓
- 警告を軽視
- ↓
- 航海を続行
- ↓
- 氷山を発見
- ↓
- 減速操作
- ↓
- 方向転換操作
- ↓
- 衝突

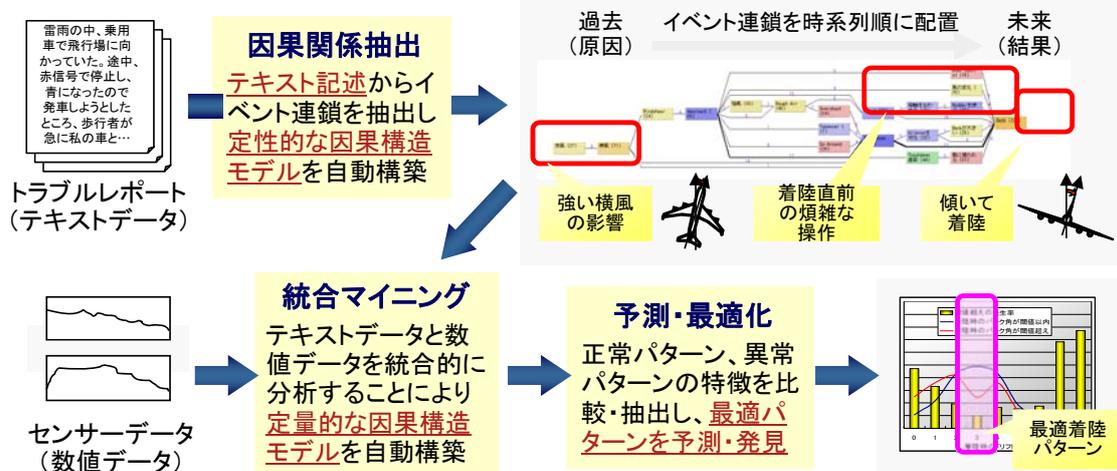
抽出されたリスクシナリオの例

交通事故レポート(500件)のイベント連鎖モデル



テキストデータと数値データの統合

- テキスト+数値データから**定量的な因果モデル**を自動構築
→ 経産省「情報大航海プロジェクト」にて、JALと実証実験を実施



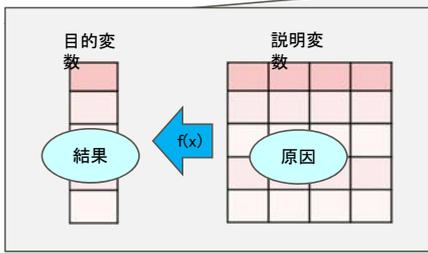
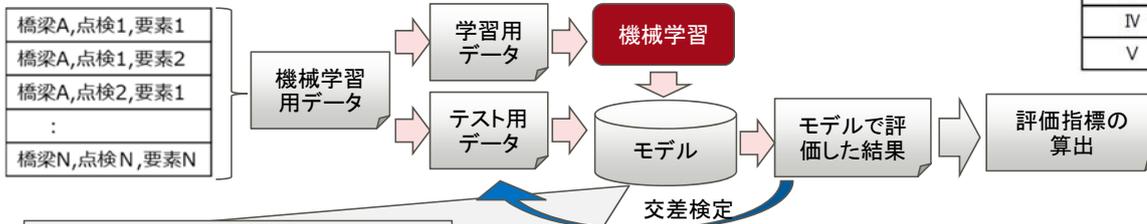
統合分析(テキスト+センサー)により強い横風での最適着陸パターンを発見

劣化量推定／予測 (近似モデルによる説明可能なAI)

劣化量推定：橋梁の健全度の推定

- 点検結果データから健全度(変状グレード)を推定
- ベテラン技術者の暗黙知が再現できるかを検証

変状グレード	変状や劣化の進行
I	問題となる変状がない
II	軽微な変状が発生している
III	変状が発生している
IV	変状が著しい
V	深刻な変状が発生している



【説明変数】
学習するパターン(条件)となるデータ項目(損傷種別/状態/拡がり/損傷パターン、材質番号)

【目的変数】
予測対象(結果)となるデータ項目(健全度)

【交差検定】
データを学習用とテスト用に分け、それぞれ入れ替えながら評価

劣化量推定：橋梁の健全度の推定

- 人間が判断した健全度を高精度で再現
- ↓
- 点検データから人間と同等レベルの健全度判断が可能
- 学習されたモデルから、人間が行っている判断ロジックを把握することが可能

【主桁(鋼)・1刻み】

		予測値					総計
		1	2	3	4	5	
健全度	1	8	1	1			10
	2		17	82			99
	3		2	2180	275	26	2483
	4			300	15883	707	16890
	5			32	1778	32061	33871
総計		8	20	2595	17936	32794	53353

【主桁(コンクリート)・1刻み】

		予測値					総計	
		1	2	3	4	5		
健全度	1	21	7	2			30	
	2	2	84	16			102	
	3		1	554	99		657	
	4			4	53	57362	1	57420
	5				728	22	750	
総計		24	98	625	58189	23	58959	

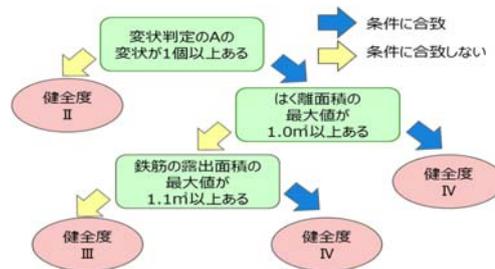
	主桁(コンクリート)		主桁(鋼)	
	平均誤差	平均絶対誤差	平均誤差	平均絶対誤差
健全度 1	0.418	0.418	0.518	0.518
健全度 2	0.292	0.346	0.809	0.814
健全度 3	0.129	0.200	0.237	0.251
健全度 4	0.009	0.013	0.089	0.131
健全度 5	-0.833	0.833	-0.064	0.064
総計	0.000	0.026	0.000	0.096

人間の健全度判断を高精度で再現

機械学習モデルにおける説明変数(点検項目)の重要度

変数名	主桁		支床(鋼)	
	コンクリート	鋼	鋼	ゴム
材質種別番号	268.583	154.148	53.268	0.000
⑤-1びび割れ(PC) 状態大	65.046	0.000		
⑤-2びび割れ(PC) 拡がり大	63.523	0.000		
⑦-1はく離・鉄筋露出(塩害・中性化) 拡がり小	114.650	0.000		
⑦-1はく離・鉄筋露出(塩害・中性化) 状態大	72.827	0.000		
⑦-1はく離・鉄筋露出(塩害・中性化) 状態中	64.086	0.000		
⑦-2はく離・鉄筋露出(塩害) 状態中	63.239	0.000		
⑧-1はく離れからの露出(水・塩害石灰・中性化) 拡がり大	86.145	0.000		
⑧-1はく離れからの露出(水・塩害石灰・中性化) 状態大	85.582	0.000		
⑧-1はく離れからの露出(水・塩害石灰・中性化) 状態中	77.830	0.000		
⑧-1はく離れからの露出(水・塩害石灰・中性化) 状態小	0.000	406.011		
⑧-1コンクリート補修材の埋置(鋼筋) 拡がり大	0.000	707.554	0.000	0.000
⑧-1コンクリート補修材の埋置(鋼筋) 状態中	0.000	3745.846	0.000	0.000
⑧-1コンクリート補修材の埋置(鋼筋) 状態小	0.000	0.000	0.000	0.000
⑧-2支床の埋置(鋼筋) 拡がり(-)	278.639	73.722		
⑧-2支床の埋置(鋼筋) 状態小	277.441	78.416		
⑧-2支床の埋置(鋼筋) 状態中	3.041	134.205		
⑧-2支床の埋置(鋼筋) 状態大	0.674	266.353		
⑧-2支床の埋置(鋼筋) 状態中	0.203	25.357		
⑧-2支床の埋置(鋼筋) 状態小	1.823	283.601		
⑧-2支床の埋置(鋼筋) 状態大	0.000	276.909		
⑧-2支床の埋置(鋼筋) 状態中	5.146	20.938	5.870	108.745
⑧-2支床の埋置(鋼筋) 状態小	25.147	131.047		
⑧-2支床の埋置(鋼筋) 状態大	5.254	21.860	25.207	132.144

決定木(ルールベースの判断ロジック)の可視化

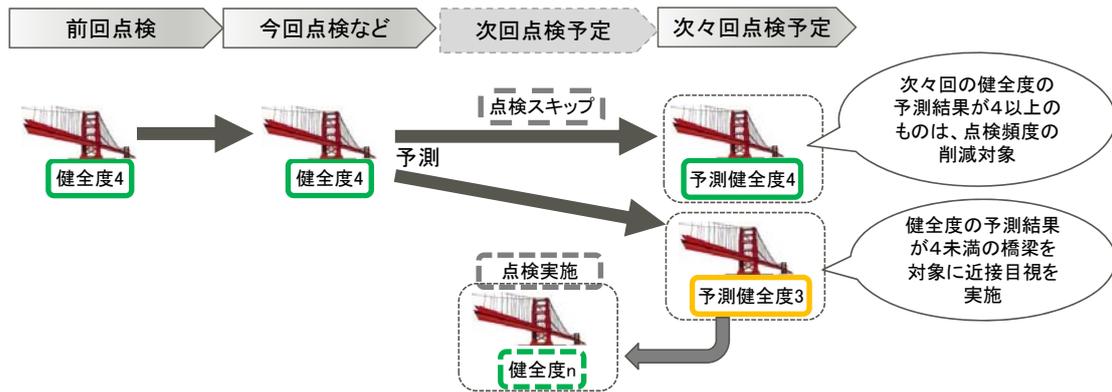


■ 分析の目的

- 劣化予測をすることで点検をスキップすることが可能かを実データを用いて検証

■ 評価結果

- 健全度の予測により(理屈の上では)点検を半減することが可能



【第四部】「信頼できる診断と予測」の実現に向けて

予知保全の課題と解決方法

カテゴリ	課題	概要	解決方法
抽象化・汎用化が不完全	個別性	対象特化したモデルの作りこみが必要	【解決方法】ハイブリッドモデル ・対象分野における 知識を活用 ・物理モデル／経験測とデータドリブンのアプローチを統合 ・Black Box + White Box
	個体差	異質の個体を混ぜると良いモデルができない＋異質の個体の予測ができない	
	状態変化	状態が時々刻々と変化(モード、負荷、…)	
	事例不足	異常事例(教師データ)が少ない＋データが不足する範囲の精度が不十分	
結果の価値・使い方	説明性	異常検知しても異常のタイプや場所が不明だと役に立たない	【期待効果】 ・抽象化／汎用化が可能 ・ 実務者視点での解釈可能性 が向上 ・適切な劣化過程を選定可能 ・対象分野を限定することでモデリング／カスタマイズ不要
	因果推定	異常の原因が不明 → 劣化過程が不明 → 劣化予測が不正確	
実務適用の課題	知識・スキル	モデル作成に高度な知識・スキルが必要	
	投資対効果	コスト効果が見積もれないと投資が困難	

予知保全のアプローチ

カテゴリ	手法	概要	例	モデルのよりどころは知識
モデルベース (モデルのパラメータはデータから推定)	物理モデル	劣化過程や因果関係を物理則に基づいて記述	<ul style="list-style-type: none"> Forman則(亀裂進展) Paris則(疲労亀裂進展) 	
	知識ベース	専門家やルールセットの情報から得られる知識を統合して活用	<ul style="list-style-type: none"> エキスパートシステム ファジロジック 	
	確率過程／統計モデル	劣化過程を確率過程／統計モデルとして記述	<ul style="list-style-type: none"> ガンマ過程、自己回帰 比例ハザード、ロジスティック回帰 	
モデルレス(データドリブン)		特定のモデルを規定しない	<ul style="list-style-type: none"> SVM、RVM、Random Forest、DL 	

■ これまで

モデルのよりどころはデータ

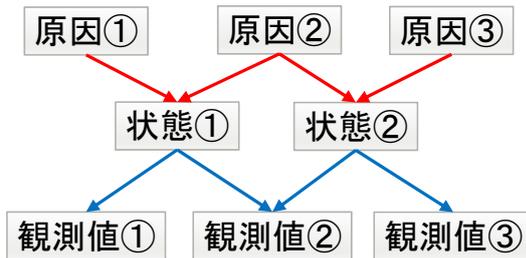
- 物理モデル／知識ベースによる**専門家向けの方法論** → 振動解析による故障診断
- **単独モデル** → 物理モデル or 確率過程／統計モデル or モデルレスの手法

■ 今後:ハイブリッドアプローチ(解釈可能なモデルベース＋高精度なモデルレス)

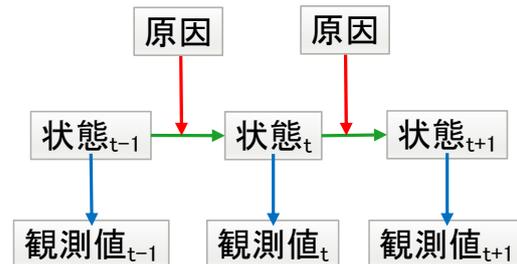
- 物理モデル＋知識ベース＋確率過程／統計モデル＋モデルレス

モデル	概要	活用プロセス	例
現象／観測モデル 状態と観測値	各状態はどの観測値にどのように現れるか	<ul style="list-style-type: none"> 状態推定 異常検知・分類 	<ul style="list-style-type: none"> 振動工学 回転機械の振動診断
因果モデル 原因と状態(変化)	状態変化は何によって引き起こされるか	<ul style="list-style-type: none"> 原因推定 劣化モデル選定 	<ul style="list-style-type: none"> コンクリート構造物の劣化原因(塩害、凍害、中性化)
劣化モデル 状態の変化	状態はどのように変化するか	<ul style="list-style-type: none"> 状態予測 余寿命予測 	<ul style="list-style-type: none"> 故障物理モデル

【静的なとらえ方】



【動的なとらえ方】



「信頼できる診断と予測」の実現に向けて

■ 「信頼できる診断と予測」の実現に必要な技術要素

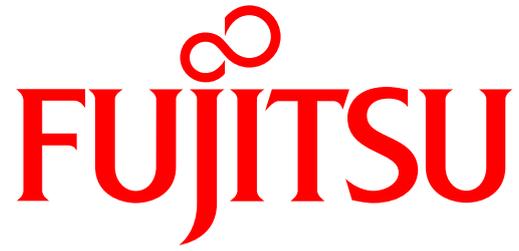
- ハイブリッドアプローチ (Black Box + White Box = Grey Box)
- 説明可能なAI／解釈可能な機械学習 + 第一原理モデル(物理モデル or 経験測)

■ Black BoxとWhite Boxのバランスやベースのモデルはタスク依存

- 異常検知(異常かどうかの二値判別): DLで高精度に検知・予測
- 異常分類(異常の程度・種類の多値判別): 物理モデル／経験則ベースの現象モデル
- 原因推定(分類や原因の多値判別): 物理モデル／経験則ベースの因果モデル
- 余寿命予測・故障予測(劣化の定量的予測): 物理モデル+確率過程／統計モデル

■ 実現に向けた技術課題

- 物理モデルや経験則の抽出・モデリング
- モデルパラメータの動的な推定 → 状態空間モデル、デジタルツイン
- モデルの統合方法(ハイブリッド) → ベイジアンネットワーク(知識・因果+確率)



shaping tomorrow with you