



製造業におけるAI+ビッグデータ 活用の方および課題

青山学院大学 理工学部 経営システム工学科
小野田 崇

onoda@ise.aoyama.ac.jp



講演内容

- 人工知能ブームの歴史
 - 機械学習研究の歴史
- 統計学と機械学習
 - その類似性と相違
- 事例に見るAI+データの利活用
 - 設備診断における事例
- AI+データの利活用
 - 利活用には何が重要か

講演内容

- 人工知能ブームの歴史
 - 機械学習研究の歴史
- 統計学と機械学習
 - その類似性と相違
- 事例に見るAI+データの利活用
 - 設備診断における事例
- AI+データの利活用
 - 利活用には何が重要か

人工知能ブーム

- 第1次人工知能ブーム
 - 1950年代：「Artificial Intelligence(人工知能)」という呼び名が産声を上げる
- 第2次人工知能ブーム
 - 1980年代：第5世代コンピュータプロジェクト, エキスパートシステム
- 第3次人工知能ブーム
 - 2010年代：ビッグデータ + 深層学習

第1次人工知能ブーム

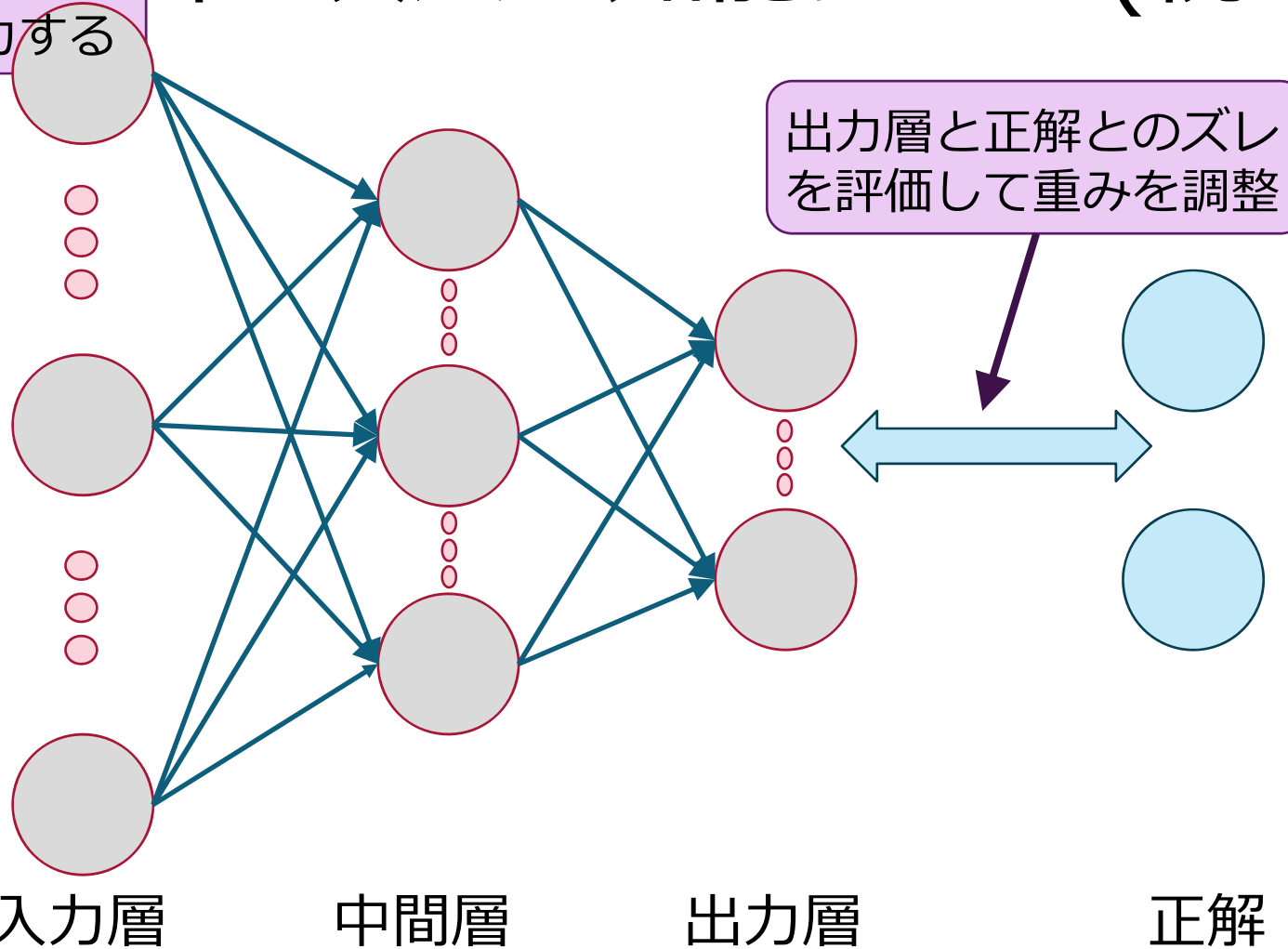
- 1950年代
 - アラン・チューリングとクロード・シャノンが「コンピュータにもチェスを指すことが可能で、原理的には世界チャンピオンにも勝てる」と発言
- 1956年
 - 若手の研究者が集結したダートマス会議において、ジョン・マッカーシーという研究者によって人工知能(Artificial Intelligence: AI)という名前が産声
- 1950年代の人工知能
 - チェスを指すコンピュータ, 数学の定理証明をするコンピュータ **パーセプトロンの提案**

第2次人工知能ブーム(1980年代)

- エキスパートシステム
 - スタンフォード大の感染症診断治療支援エキスパートシステム (MYCIN) の登場
 - 大手企業を中心に人工知能関連部署を新設
- 第五世代コンピュータプロジェクト
 - 通産省が550億円をかけて行ったプロジェクト
 - 世界が追随
 - 一階述語論理に基づいてデータ間の関係を示す命題として記述：Prolog
 - 一階述語論理:個体の量化のみを許す述語論理
 - 量化：論理式が適用される議論領域の個体の「量」の指定
 - 述語論理：数理論理学における記号的形式体系群を指す用語

第2次人工知能ブーム(続き)

学習させたい特
徴量を入力する



第2次人工知能ブーム(続き)

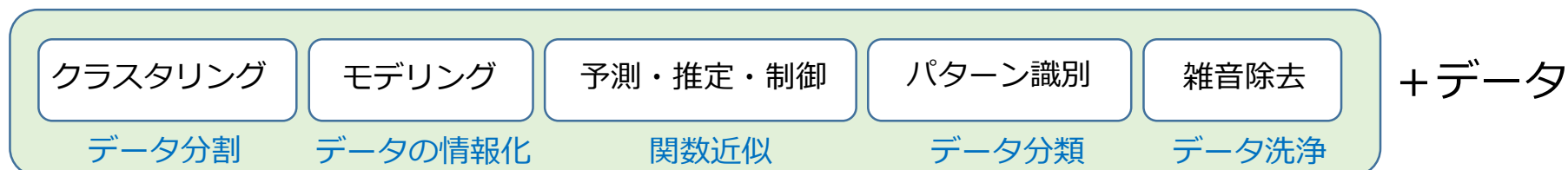
- パーセプトロン限界説への反発
- ニューラルネットワーク
 - 人間の脳神経回路を真似ることによってデータを人間と同じように分類しようというアイデア
 - 誤差逆伝播学習法
 - ある程度複雑な関係を学習：様々な産業分野に適用
 - 3層程度のシミュレーションしかできなかったため、より人間に近い、人間を超える性能を発揮するという視点からは限界
- 知識獲得≡機械学習

裏人工知能≡機械学習

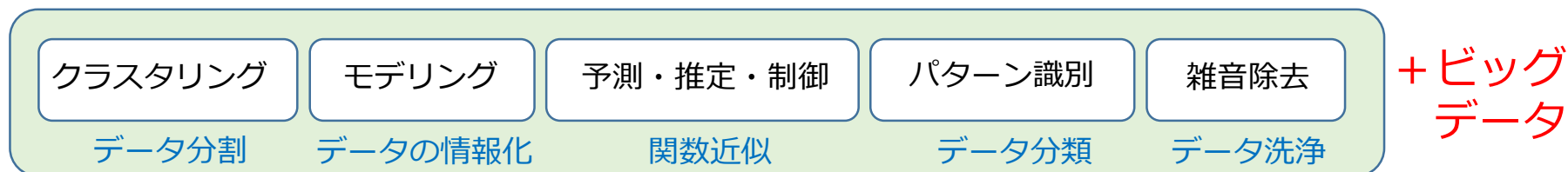
- 1990年代～2000年代初頭
- 1990年代：機械学習の成長期
 - サポートベクトルマシン, カーネル法, アンサンブル学習(Bagging, AdaBoost), 汎化理論(機械学習の性能評価)
- 2000年代：機械学習の発展期
 - ベイズモデル(変分ベイズ, Gibbsアルゴリズム), 転移学習, 半教師有り学習, 深層学習(最初に提案された時代)

第3次人工知能ブーム

• 2010年代（現在）



～2010年代：機械学習



2010年代～：人工知能(AI)

※図中うすい緑四角部分は機械学習が主に扱っている適用領域

ALPHAGOに関するニュース

- AlphaGo以前
 - 日本の囲碁プログラムZen等はプロ棋士に4子局で勝利
 - 人間のチャンピオンレベルになるのは10年後くらいと思われていた
- 2015年10月
 - AlphaGoがヨーロッパチャンピオンに5勝0敗
- 2016年1月
 - Google Deep Mind がNature 誌にAlphaGoに関する論文を発表
 - 市販ソフト(Zen, Crazy, Stone, …)に対し494勝1敗

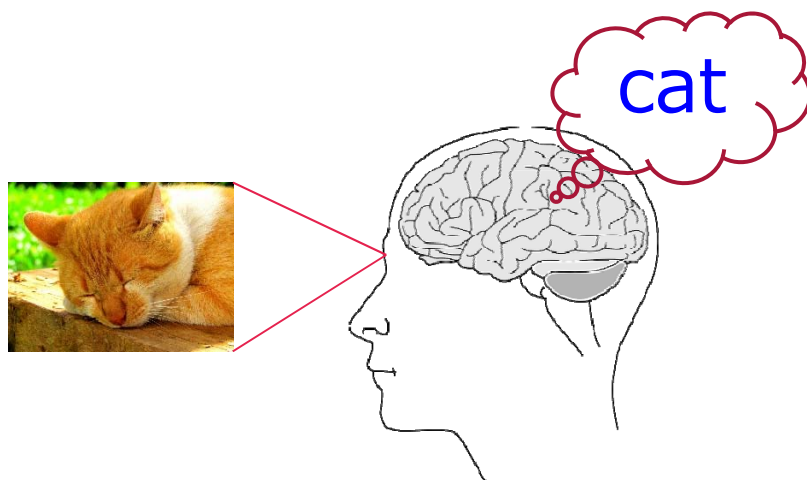
ALPHAGOに関するニュース

- 2016年3月9-15日 Google DeepMind Challenge Matchにて世界チャンピオンに4勝1敗
 - 日程：3月9日(水)-15日(火)
 - 会場：韓国・ソウル市 フォーシーズンズホテル ソウル
 - 賞金：100万ドル(約1億1千万円)
 - 持ち時間：持ち時間2時間/60秒の秒読み3回

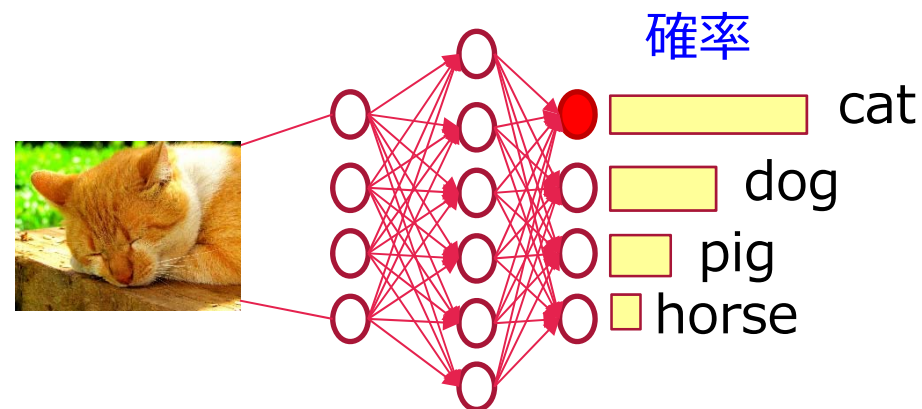
深層学習とは？

- ニューラルネットワークによる機械学習手法
- 事例：2012年一般画像認識コンペで圧勝

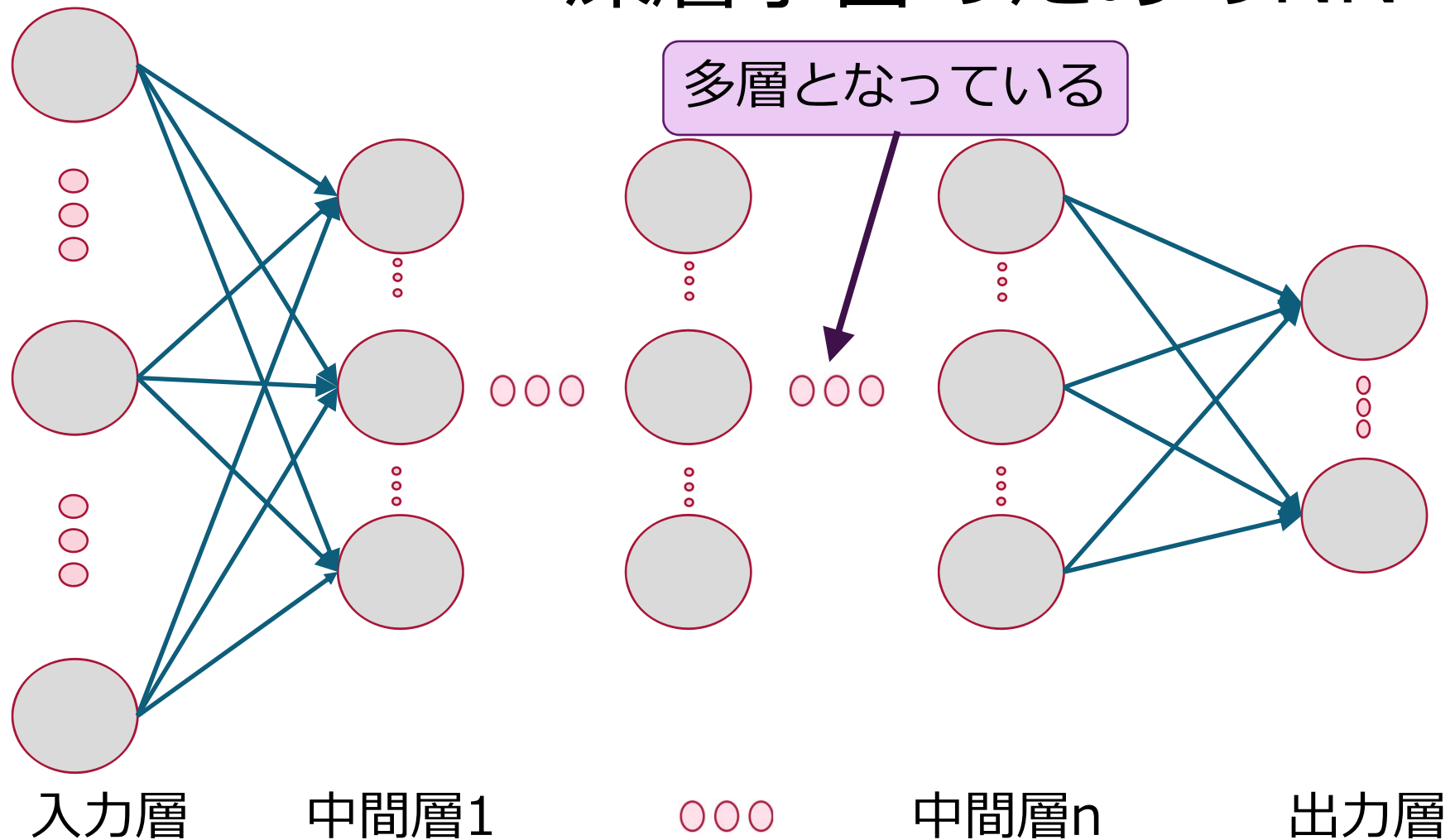
人間の知覚



機械学習(深層学習)

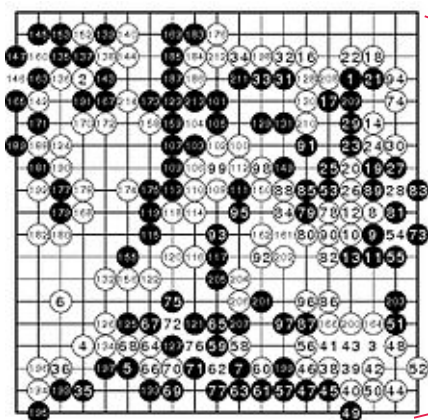


深層学習のためのNN

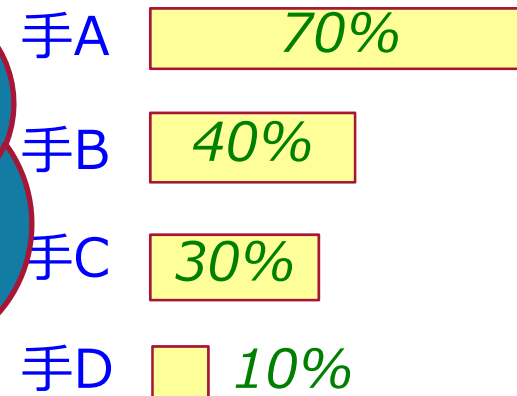


囲碁における深層学習

- どんな盤面を入力しても，プロみたいな手が打てる何か？



12層もある深いネットワーク



- このようなネットワークでプロの棋譜を学習して作る！

ICML 2017

- 44 Google, 33 Microsoft, 32 CMU
- 25 DeepMind, 23 MIT, 22 Berkeley
- 22 Stanford, 16 Cambridge, 16 Princeton
- 15 None, 14 Georgia Tech, 13 Oxford
- 11 UT Austin, 10 Duke, 10 Facebook
- 9 ETH Zurich, 9 EPFL, 8 Columbia
- 8 Harvard, 8 Michigan, 7 UCSD

講演内容

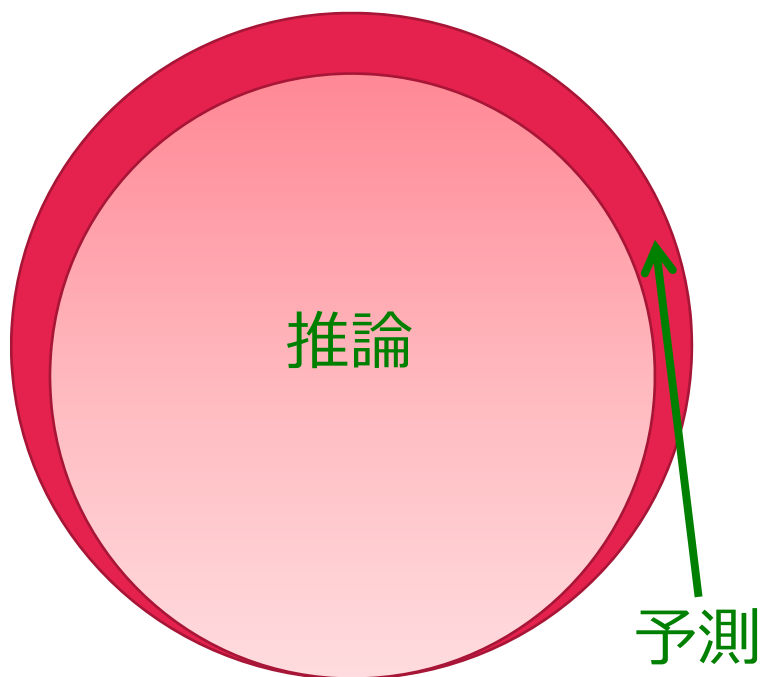
- 人工知能ブームの歴史
 - 機械学習研究の歴史
- 統計学と機械学習
 - その類似性と相違
- 事例に見るAI+データの利活用
 - 設備診断における事例
- AI+データの利活用
 - 利活用には何が重要か

統計学と機械学習(1)

- 統計学
 - 確率・統計
 - 多変量解析
 - 主に, 回帰分析, 判別分析, クラスタ分析
 - 最適化手法
 - モデル化手法
- 機械学習
 - 統計的機械学習 ← 確率・統計, 最適化手法, モデル化手法
 - パターン識別 ← 判別分析, 最適化手法
 - 非線形関数近似 ← 回帰分析, モデル化手法, 最適化手法
 - クラスタリング ← クラスタ分析, 最適化手法

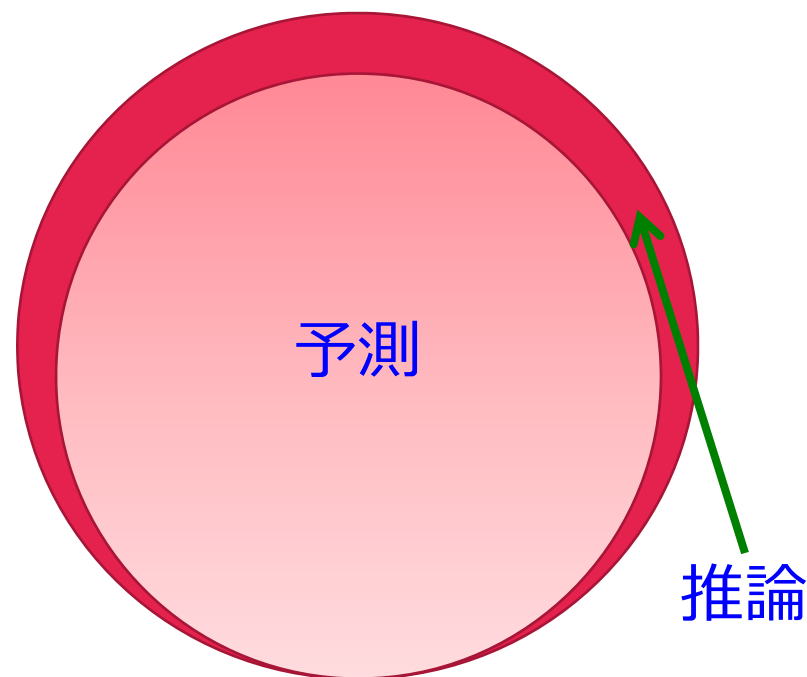
統計学と機械学習(2)

統計学




推論(分析)中心

機械学習



予測中心


機械学習とは？


$y = f(x; \theta)$
 Input x  y Output (Target value)

汎化誤差 (期待損失) $E\{L\} = \int \underline{L(y, f(x))} p(x, y) dx dy$

損失関数

損失関数の例

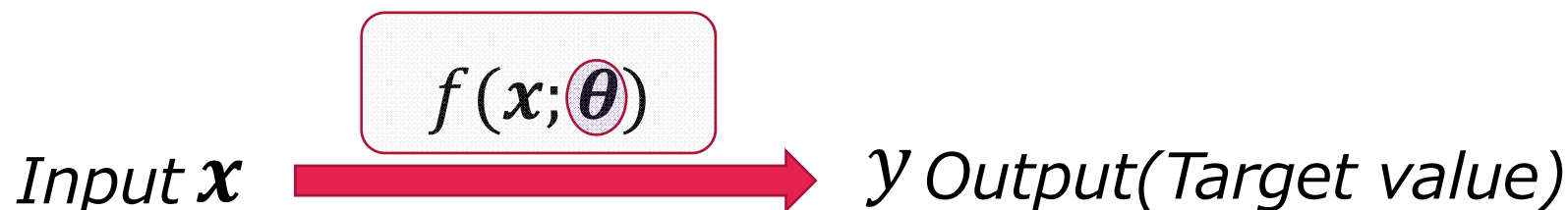
$L(y, f(x)) = \|y - f(x)\|^2$ 自乗誤差  最小自乗法

$L(y, f(x)) = -\log p(x, y)$ 対数損失  最尤推定法

統計学と機械学習

統計学は入力 x と出力 y とがどのようなモデルで説明できるかを探る！

統計学ではモデル $f(x; \theta)$ が重要！
統計学では特にパラメータ θ が重要！



機械学習では $f(x; \theta) = y$ が重要！

機械学習ではモデル $f(x; \theta)$ は重要ではない！
機械学習ではパラメータ θ は重要ではない！

機械学習は入力 x から出力 y を正確に予測するモデル探る！

講演内容

- 人工知能ブームの歴史
 - 機械学習研究の歴史
- 統計学と機械学習
 - その類似性と相違
- 事例に見るAI+データの利活用
 - 設備診断における事例
- AI+データの利活用
 - 利活用には何が重要か

製造業(WIKIPEDIA)

- 原材料などを加工することによって製品を生産・提供する産業
- 家電，自動車といった工業製品から，コンビニエンスストアで売られる弁当や飲料を作る産業までを製造業



様々な部署でビッグデータを収集可能

部署あるいは現場単位でビッグデータの利用目的が異なる

製造業(国税庁)

- 食料品製造業(1), 飲料・たばこ・飼料製造業(2), 繊維工業(3), 木材・木製品製造業(家具を除く)(4), 家具・装備品製造業(5), パルプ・紙・紙加工品製造業(6), 印刷・同関連業(7), 化学工業(8), 石油製品・石炭製品製造業(9), プラスチック製品製造業(別掲を除く)(10), ゴム製品製造業(11), なめし革・同製品・毛皮製造業(12), 窯業・土石製品製造業(13), 鉄鋼業(14), 非鉄金属製造業(15), 金属製品製造業(16), はん用機械器具製造業(17), 生産用機械器具製造業(18), 業務用機械器具製造業(19), 電子部品・デバイス・電子回路製造業(20), 電気機械器具製造業(21), 情報通信機械器具製造業(22), 輸送用機械器具製造業(23), その他の製造業(25)

輸送用機械器具製造業

- 自動車・同附属品製造業
- 鉄道車両・同部分品製造業
- 船舶製造・修理業、舶用機関製造業
- 航空機・同附属品製造業
- 産業用運搬車両・同部分品・附属品製造業
- その他の輸送用機械器具製造業

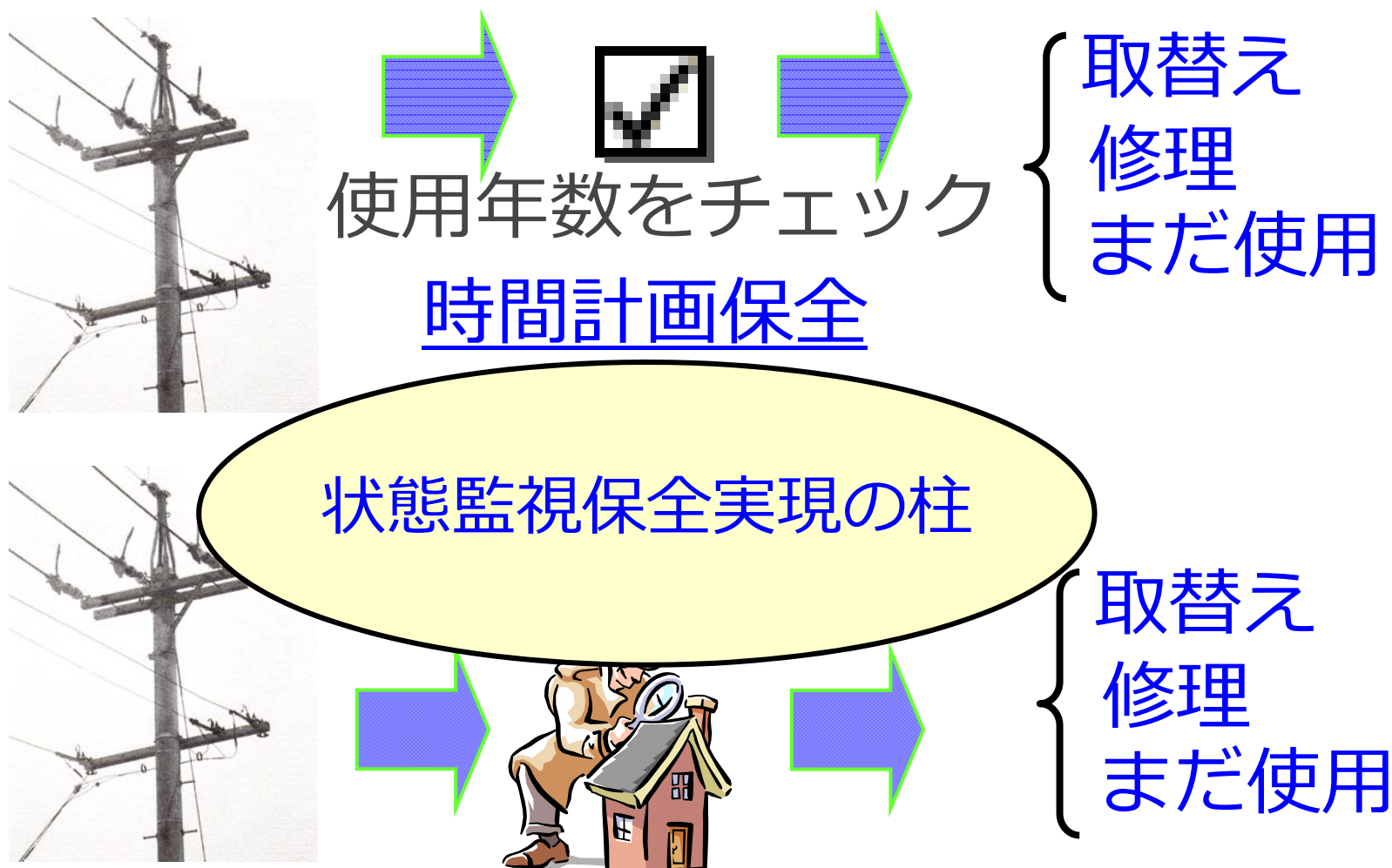
油中ガス分析データに基づく 変圧器の内部不具合診断

1. 研究の背景
2. 油中ガス分析データに基づく変圧器の内部不具合診断の現状
3. 多変量解析, 機械学習を用いた変圧器内部不具合診断
4. 実験結果
5. まとめ

油中ガス分析データに基づく 変圧器の内部不具合診断

1. 研究の背景
2. 油中ガス分析データに基づく変圧器の内部不具合診断の現状
3. 多変量解析，機械学習を用いた変圧器内部不具合診断
4. 実験結果
5. まとめ

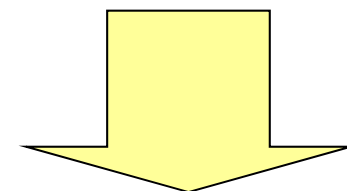
機器診断技術の必要性の変化



電力用変圧器



対象変圧器は
油入変圧器



油内のガスの
成分により内
部不具合わか
る？

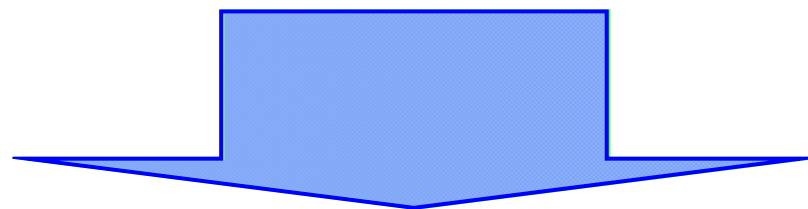
東芝産業機器製造株式会社カタログより抜粋

油中ガス分析データに基づく 変圧器の内部不具合診断

1. 研究の背景
2. 油中ガス分析データに基づく変圧器の内部不具合診断の現状
3. 多変量解析, 機械学習を用いた変圧器内部不具合診断
4. 実験結果
5. まとめ

電力用変圧器良否判定の現状

- 電気協同研究会が平成11年2月に示した判定基準を採用
- 当時より精密な油中ガス分析を実施
- 精密な油中ガス分析データを蓄積

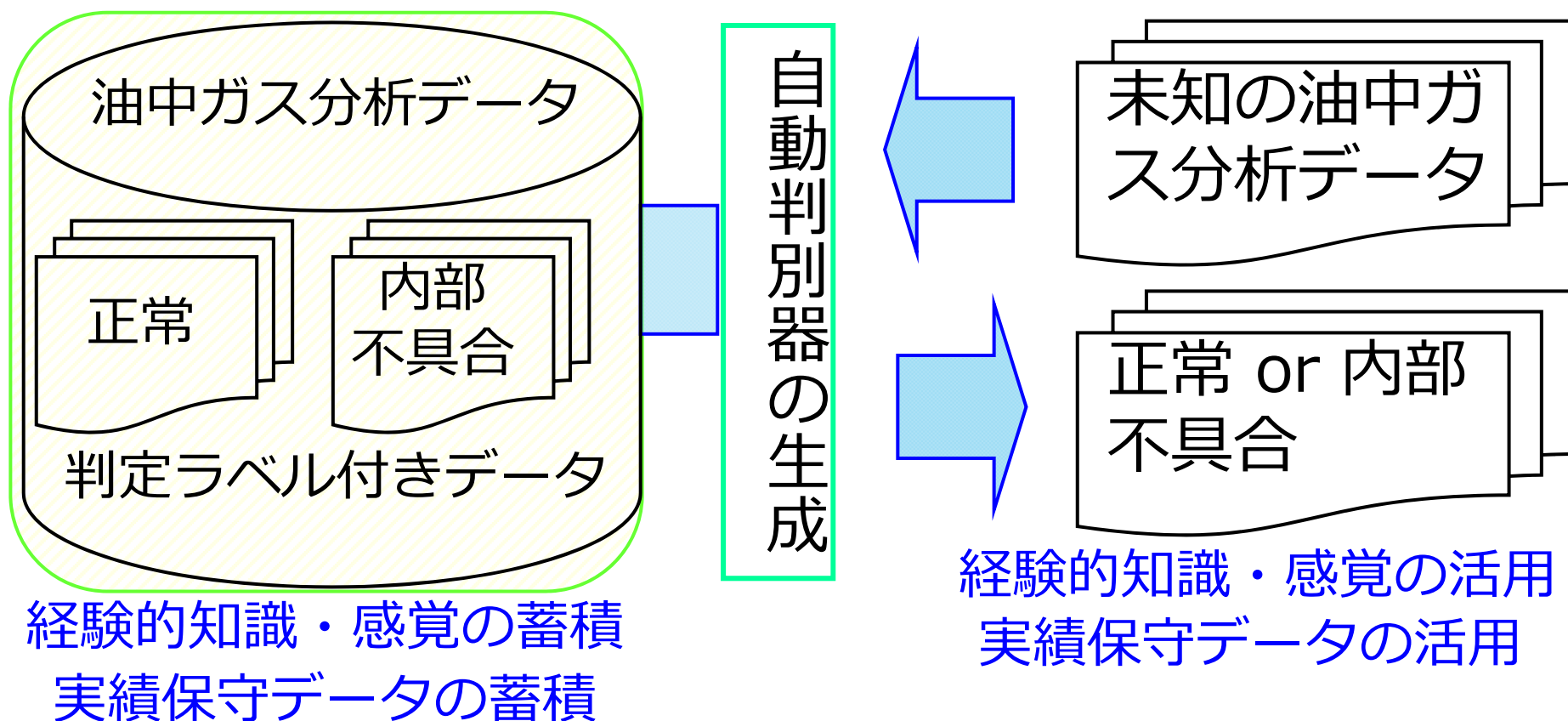


データに基づく基準の作成

油中ガス分析データに基づく 変圧器の内部不具合診断

1. 研究の背景
2. 油中ガス分析データに基づく変圧器の内部不具合診断の現状
3. 多変量解析, 機械学習を用いた変圧器内部不具合診断
4. 実験結果
5. まとめ

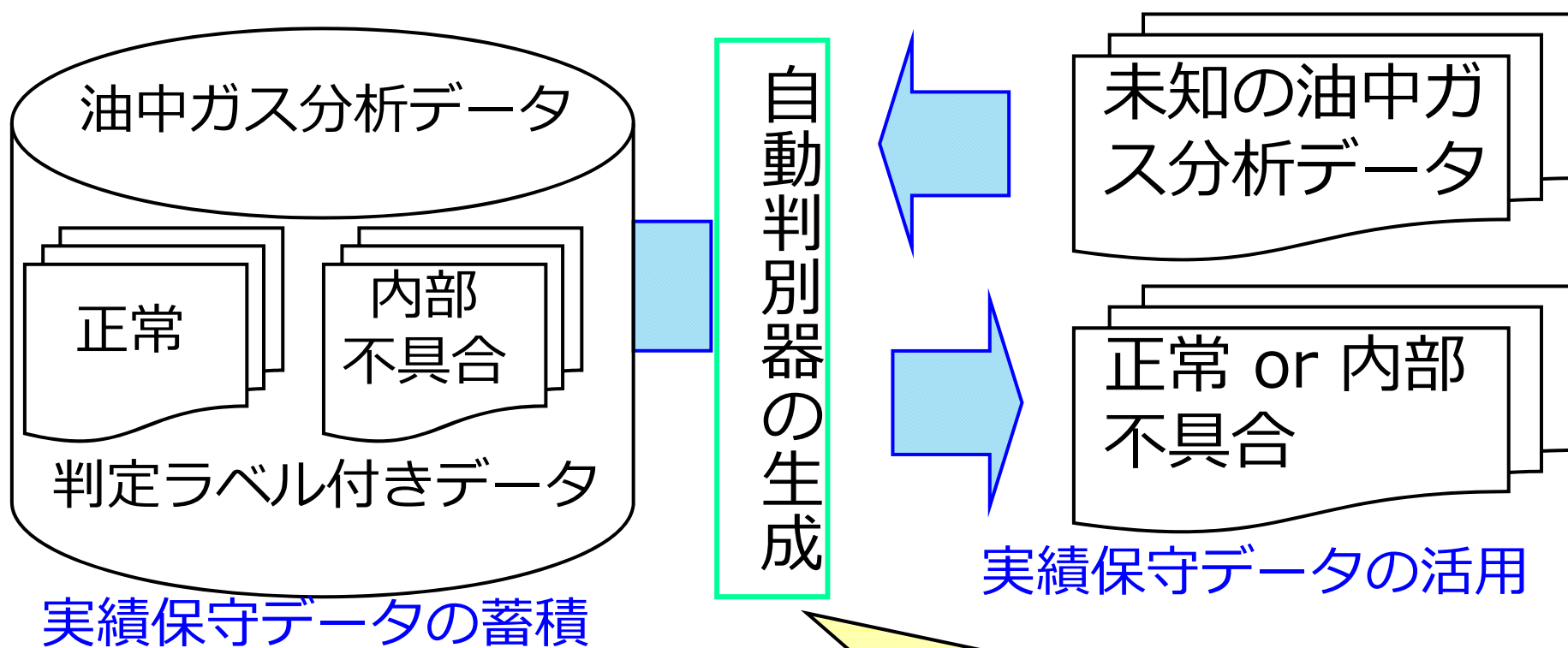
データに基づく判別(1)



油中ガス分析データ(ある電力会社)

- 115件の変圧器の油中ガス分析データ
 - 内部不具合変圧器15件, 正常状態変圧器100件
- 油中ガス分析データは7種類のガス構成
 - 水素(H₂), メタン(CH₄), エタン(C₂H₆), エチレン(C₂H₄), アセチレン(C₂H₂), 一酸化炭素(CO), 二酸化炭素(CO₂)の発生量(単位:ppm)
- 正常・内部不具合は変圧器を分解点検し目視により判定

データに基づく判別(2)

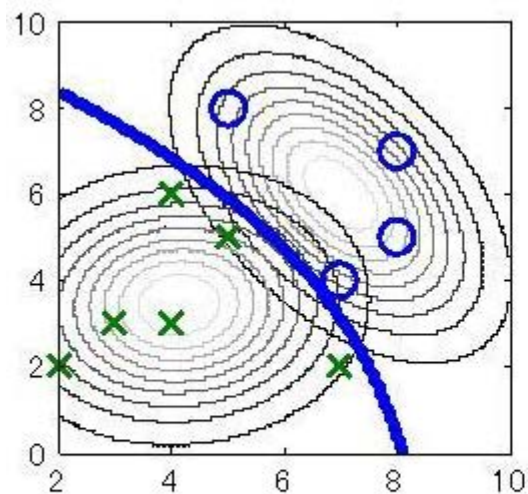


多変量解析, 機械学習による判別器の生成

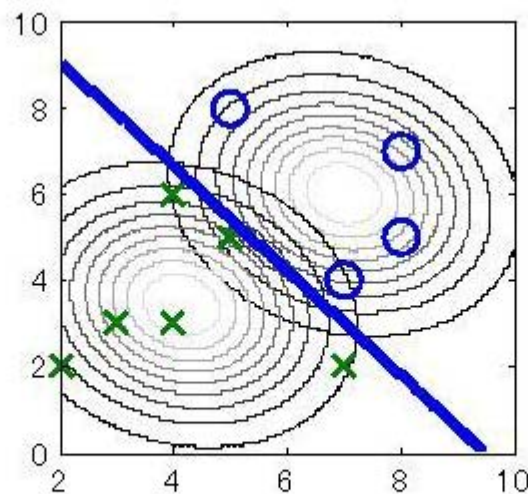
油中ガス分析データに基づく 変圧器の内部不具合診断

1. 研究の背景
2. 油中ガス分析データに基づく変圧器の内部不具合診断の現状
3. 多変量解析，機械学習を用いた変圧器内部不具合診断
4. 実験結果
5. まとめ

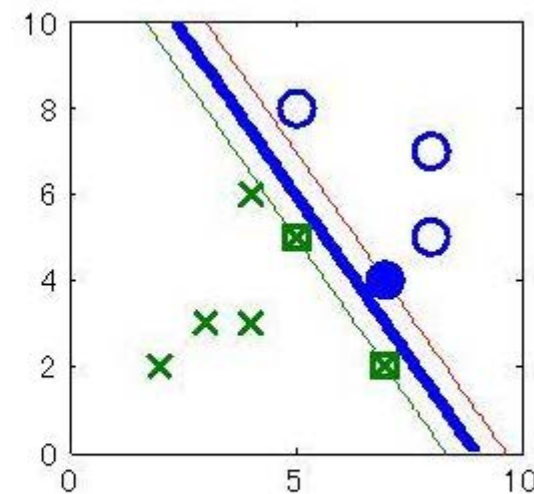
代表的な多変量解析



(a) マハラノビス距離



(b) 多変量線形判別



(c) 線形SVM

- マハラノビス距離：正規分布を仮定
- 多変量線形判別：正規分布，同一分散共分散を仮定
- 線形SVM：必ず識別できることを仮定

マハラノビス距離の適用結果

- データは正規分布していると仮定
- 与えられたデータに対し、内部不具合変圧器を正常変圧器と誤判別する場合はなかった。
- 100件の正常変圧器の内、50件を内部不具合変圧器と誤判別した。
- 与えられたデータでさえ正確な判別を実現できなかった。

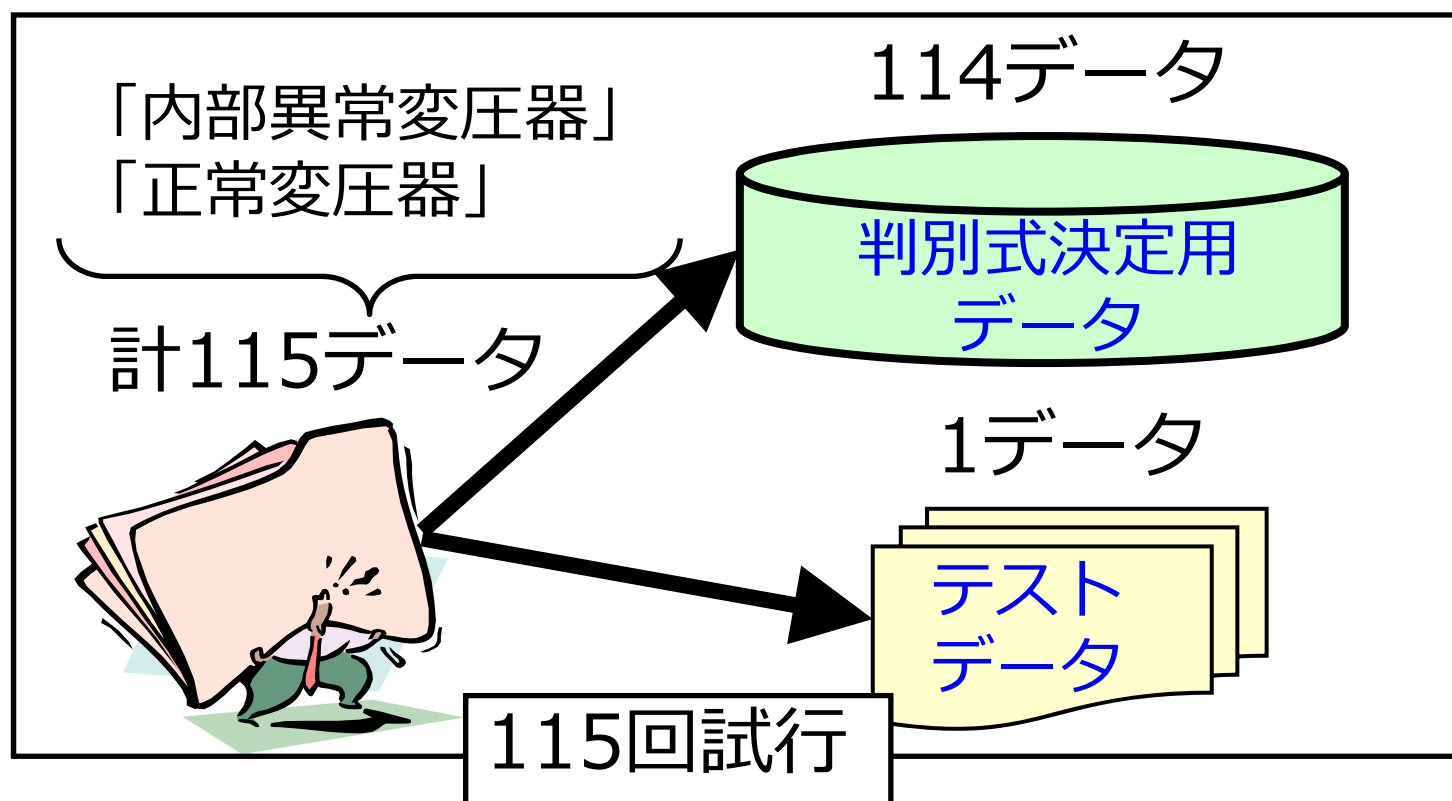
線形判別分析適用結果

- 与えられたデータに対し、正常変圧器を内部不具合変圧器と誤判別する場合はなかった。
- 内部不具合変圧器を正常変圧器と誤判別する場合は6件(15件中)存在した。
- 誤判別した6件はいずれも C_2H_4 (エチレン) または C_2H_2 (アセチレン) の発生量が小さいものだった。

SVM適用結果

- 与えられたデータに対し、正常変圧器を内部不具合変圧器と誤判別する場合はなく、内部不具合変圧器を正常変圧器と誤判別する場合もなかった。
- LOO(Leave-One-Out)検定を行った結果、平均誤判別率は0%であった。

LEAVE-ONE-OUT



SVMによる説明変数の削減

正規化	あり	なし
H ₂ (水素)	-0.6079	
CH ₄ (メタン)	-5.0958	
C ₂ H ₆ (エタン)	0.9201	
C ₂ H ₄ (エチレン)	-5.9970	- 0.0201
C ₂ H ₂ (アセチレン)	-23.4126	- 0.3332
CO (一酸化炭素)	-0.0151	
CO ₂ (二酸化炭素)	0.6453	
切片	1.1744	1.1342

2種のガスデータによる判別

$$Z = -0.0201 \times C_2H_4 - 0.3332 \times C_2H_2 + 1.1342$$

各ガスの部分にはppmを単位としたガス量が入る

Z が正なら正常変圧器, さもないければ
内部不具合変圧器

- 与えられたデータを全て正しく判別(識別).
- LOOによる平均誤識別率の評価も0%.

油中ガス分析データに基づく 変圧器の内部不具合診断

1. 研究の背景
2. 油中ガス分析データに基づく変圧器の内部不具合診断の現状
3. 多変量解析, 機械学習を用いた変圧器内部不具合診断
4. 実験結果
5. まとめ

まとめ

- マハラノビスの距離，多変量線形判別，SVMを油中ガス分析データに適用した。
 - 現状の油中ガス分析データの収集状況から，多変量線形判別分析，マハラノビスの距離を用いた変圧器の状態判別の実現は，**困難**であることを確認
 - SVM を用いた場合，現在収集されたデータから変圧器の状態を正しく判別できる線形判別式を生成できる可能性を確認
- 様相診断へのSVMの適用可能性の検討

講演内容

- 人工知能ブームの歴史
 - 機械学習研究の歴史
- 統計学と機械学習
 - その類似性と相違
- 事例に見るAI+データの利活用
 - 設備診断における事例
- AI+データの利活用
 - 利活用には何が重要か

ビッグデータの利活用に何が重要か

- 明確な利用目的
- 直感に合う

- クリティカルな課題(理由が必要な課題)への本格的な適用は？
- SE4ML: Software Engineering for Machine Learning (NIPS 2014 Workshop)
 - Machine Learning: The High Interest Credit Card of Technical Debt



ご清聴ありがとうございました

onoda@ise.aoyama.ac.jp

本著作物は原著作者の許可を得て、株式会社日本科学技術研修所（以下弊社）が掲載しています。本著作物の著作権については、制作した原著作者に帰属します。

原著作者および弊社の許可なく営利・非営利・イントラネットを問わず、本著作物の複製・転用・販売等を禁止します。

所属および役職等は、公開当時のものです。

■公開資料ページ

弊社ウェブページで各種資料をご覧いただけます <http://www.i-juse.co.jp/statistics/jirei/>

■お問い合わせ先

(株)日科技研 数理事業部 パッケージサポート係 <http://www.i-juse.co.jp/statistics/support/contact.html>