

不適合納品物判定（品質管理） へのAIの適用

- (1) AIプロジェクトチーム, (2) 構造安全評価系,
(3) 構造基盤技術系, (4) 知識・データシステム系
*馬 沖^{1,2,4)}, 平方 勝^{1,3,4)}, 小沢 匠³⁾, 間島 隆博^{1,4)}

- 背景・目的
- 使用した機械学習の種類
- 学習モデルの構築
 - 入力項目の定量化
 - 過学習の防止
 - 入力項目の重要度評価(感度解析)
- 各学習方法の予測結果
 - 各学習方法の最適解
 - 各機械学習での検査件数に不適合発見率の推移
 - 3つ機械学習モデルの相関性・妥当性の検討
- 結言

- 船舶に使用される様々な艀装品やその部品に対して検査が行われている。
 - 膨大な数のため、検査前に不適合品を予測することにより業務効率の改善が期待できる。
- 計算機の処理能力向上などを背景に、機械学習が近年着目されている。
 - 機械学習はデータ間の関連性を統計処理により抽出する手法であり、応用範囲が広い。

- 
- ✓ 実際の検査で得た各入力項目とその検査の結果(37万件)を用い、**検査の適合/不適合**を予測する機械学習モデルを構築
 - ✓ 構築したモデルの精度評価を実施

本研究では3つの機械学習方法でそれぞれ予測モデルを構築

※アルゴリズムはPythonのライブラリ関数を使用

ニューラルネットワーク(NN)

- 多層の関数によるモデル
- 繰り返し計算(学習)により, 各ユニット間の係数等を最適化

サポートベクターマシン(SVM)

定量化した入力項目を変数とした識別関数によりデータを分類

ランダムフォレスト(RF)

ランダムに抽出したデータから作成した複数の決定木から構成される予測モデル

入力項目	種類	入力項目	種類	入力項目	種類
取引先	C	初回回答納期	Db	発注累積個数	V
機種	C	回答納期	Db	資本金	Vb
装置	C	単価	V	従業員数	Vb
細分の有無	V	発注金額	V	評点	Vb
素材の有無	V	発注LT	V	代表者生年	Vb
注文日	D	部品LT	V	不適合累積件数	V
納期	D	検査区分	C	不適合累積金額	V
注文数	V	初回品マーク	V		
取消数	V	発注累積回数	V		

C: 文字データ

D: 日付データ

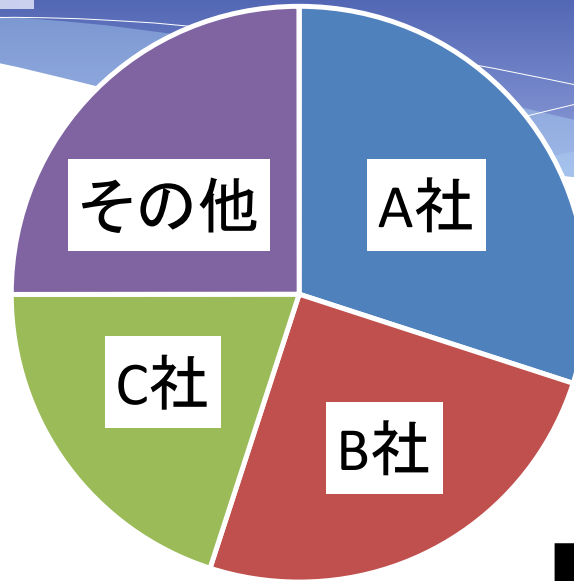
V: 数量データ

b: データの欠落あり

※ 上記入力項目+検査の適合/不適合がまとめられた**37万件**のデータ(内**0.5%が不適合**)を用いて予測モデルを構築

上記25項目を全て定量化

- 文字データ（取引先など）は構成割合の高いデータ（1%以上）に着目し、その数だけ新たに変数を設定
- そのデータに該当するか否かを1/0で標記



ID	取引先
1	A社
2	B社
3	C社
4	Z社
:	:



ID	取引先: A社	取引先: B社	取引先: C社
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0
:	:	:	:

※欠落データの割合が高ければ
欠落データの項目を追加する

日付データ（納期など）は月のみに着目して、文字データと同様の処理で定量化

ID	納期	ID	納期: 1月	同: 2月	同: 3月	...
1	2001/1/7	1	1	0	0	...
2	1998/2/6	2	0	1	0	...
3	2017/3/3	3	0	0	1	...
:	:	:	:	:	:	:

数量データの定量化

- データの欠落がない場合はそのまま使用
- 欠落がある場合は値域内で数段階程度に区分し、その区分に該当するか否かで1/0で標記し定量化

ID	注文数	ID	~10	11~100	101~
1	201	1	0	0	1
2	50	2	0	1	0
3	2	3	1	0	0
4	(欠落)	4	0	0	0
:	:	:	:	:	:

検査結果が既知の37万件データの内、90%を教師データ、10%を検証データに分類



教師データを用いて予測モデルを構築し、検証データを用いてモデルを評価
(※検証データの**評価値**が最大となるように**機械学習パラメータ**を最適化)

評価値

$$\text{評価値} = R \times P$$

※連続値で予測するため、評価値が最大となる閾値(不適合と判断する値)も同時に求める。

R: Recall(実際に不適合であった製品の内、不適合と予測出来たものの割合)

P: Precision(不適合と予測された製品の内、実際に不適合であったものの割合)

機械学習パラメータ

NN: 使用する入力項目, 隠れ層の構成(1層or4層), 学習回数

SVM: 使用する入力項目, C 及び γ (識別関数のパラメータ)

RF: 使用する入力項目

入力項目の重要度評価(感度解析)

使用する入力項目の最適化

効率的な最適化のために各入力項目の重要度を評価

RF: Pythonのライブラリ関数を使用(高頻度・高純度に分類するものを重要と評価)

NN: 各入力項目の値を変化させた場合の予測値の変化から重要度を評価

SVM: NNのものを流用

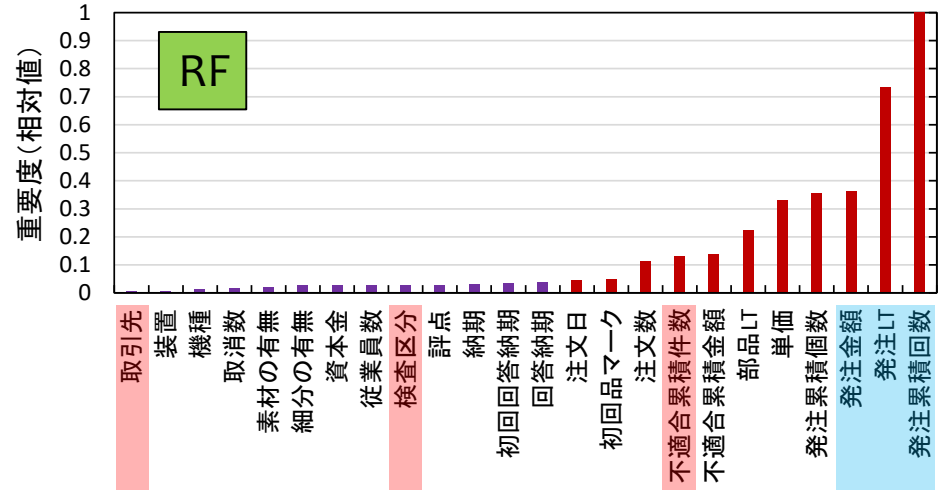
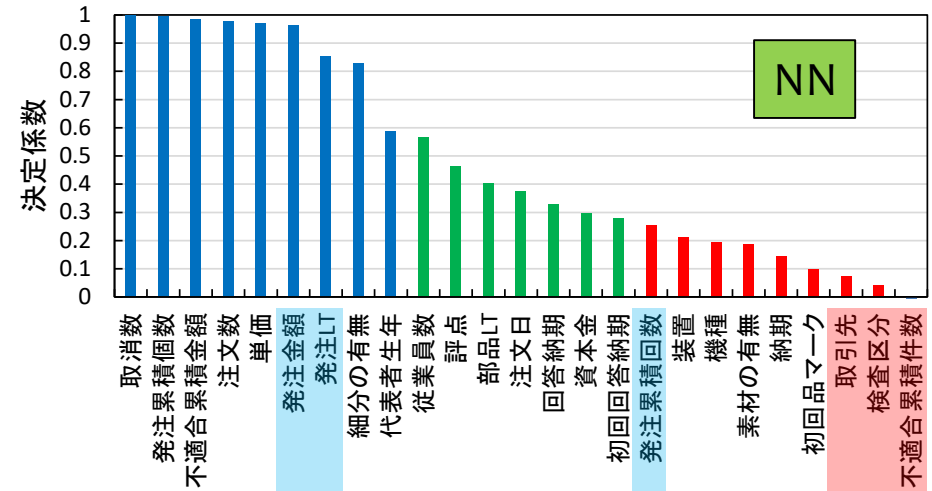
NN & SVM

Case 1	青+緑+赤
Case 2	緑+赤
Case 3	赤

RF

Case I	紫+茶
Case II	茶
Case III	紫

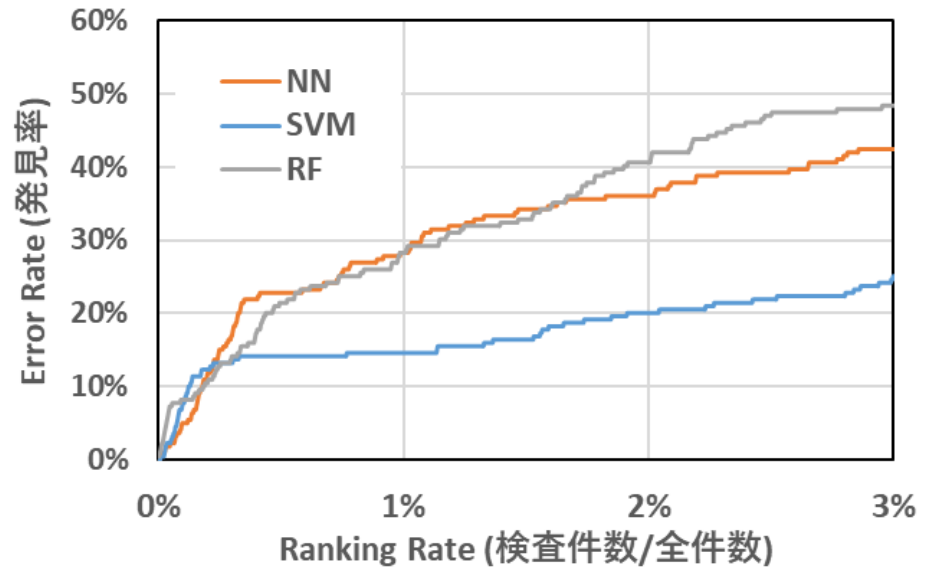
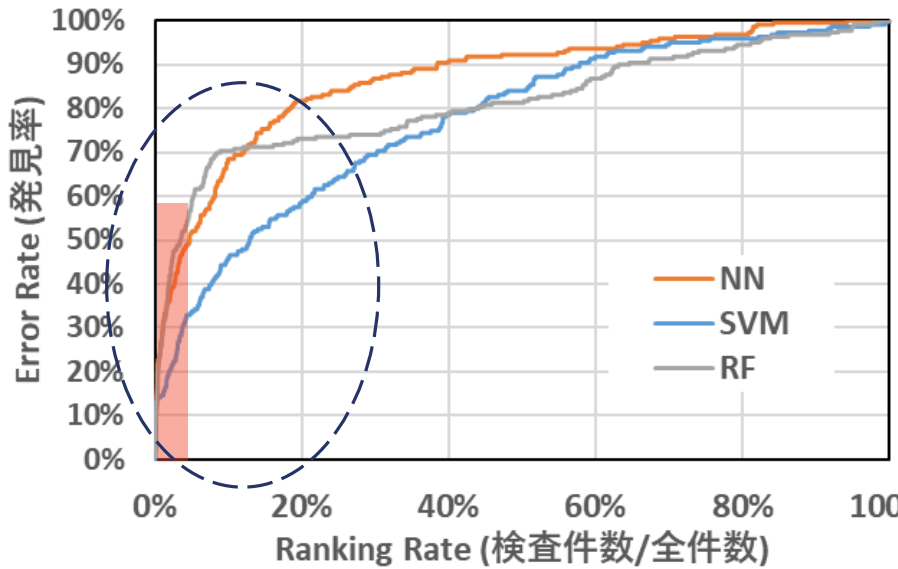
低 ← 重要度 → 高



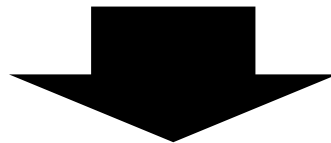
学習	評価値 (閾値)	最適パラメータ
NN	0.0803 (0.191)	入力項目:Case 3 隠れ層:4層 学習回数:400
SVM	0.0538 (-0.821)	入力項目:Case 3 C = 40, $\gamma = 0.0123$
RF	0.0688 (0.860)	入力項目:Case I

実測	不適合	不適合	適合	適合
予測	不適合	適合	不適合	適合
NN	48件 (0.128%)	171件 (0.456%)	83件 (0.221%)	37,176件 (99.195%)
SVM	24件 (0.064%)	195件 (0.520%)	27件 (0.072%)	37,232件 (99.343%)
RF	15件 (0.043%)	204件 (0.542%)	1件 (0.003%)	37,258件 (99.412%)

- NNが最も評価値が高かった
- 実際に不適合であったものを正しく予測出来た件数もNNが最も多く、48件であった。
- RFは実際に不適合であったものを正しく予測出来た件数は少ないが、実際は適合であるものを誤って不適合と予測した件数が1件のみであった。(Recallは低いものの、Precisionが高かった)



NNでは不適合の可能性の高い上位2割をデータを調べることにより、
約8割の不適合を発見可能



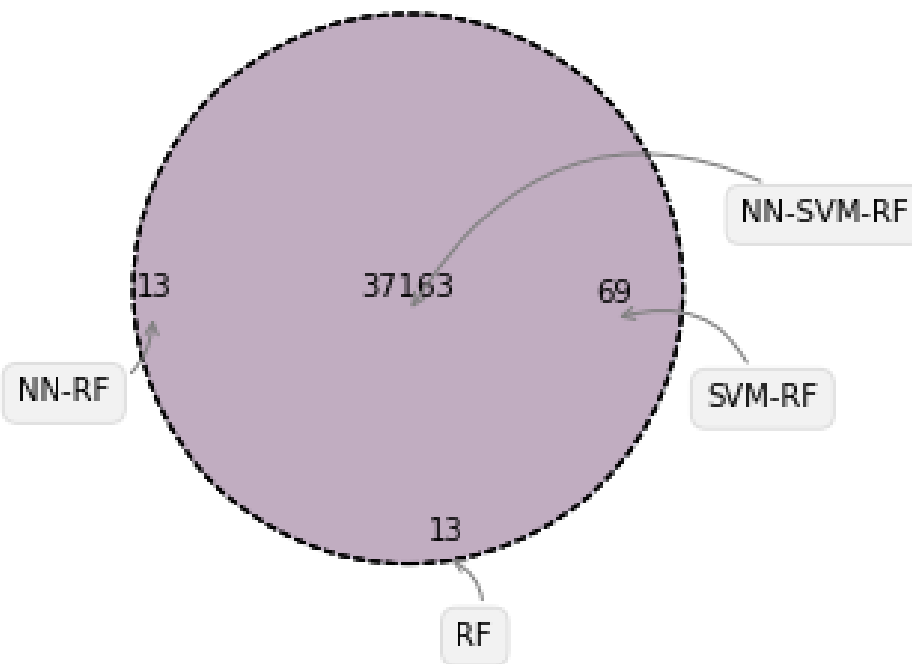
一般的な業務効率向上に十分資する

3つの**全く異なる機械学習**が満足いく結果であれば、予測結果に**相関性**があるはずである。そこで最適解に基づき各機械学習モデルの**相関性・妥当性**を検討した。

実測	不適合	不適合	適合	適合
予測	不適合	適合	不適合	適合
NN	48件 (0.128%)	171件 (0.456%)	83件 (0.221%)	37,176件 (99.195%)
SVM	24件 (0.064%)	195件 (0.520%)	27件 (0.072%)	37,232件 (99.343%)
RF	15件 (0.043%)	204件 (0.542%)	1件 (0.003%)	37,258件 (99.412%)

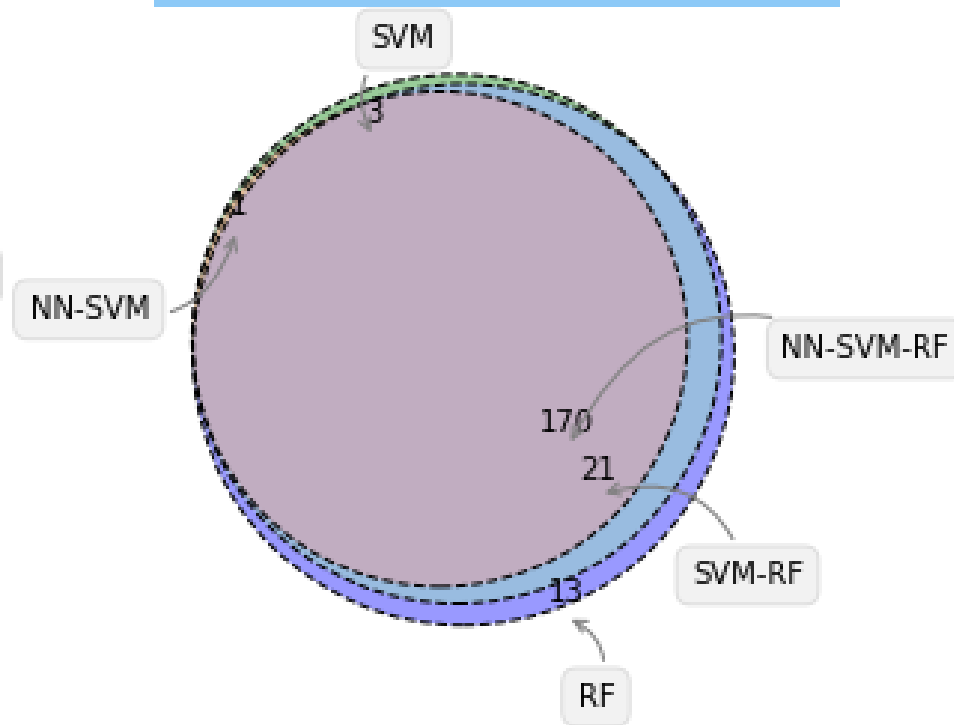
実測/予測の適合/不適合の組み合わせ**4パターン**ごとに3つの最適解で**重複したデータの割合**をそれぞれ導出

実測適合 & 予測適合



3つの機械学習モデルのデータの99.7%以上が重複

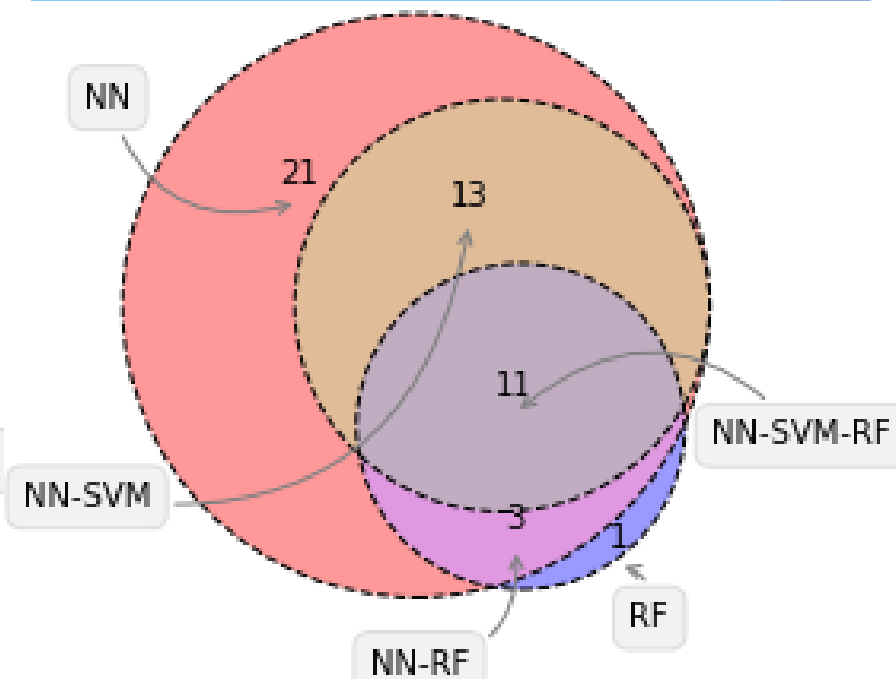
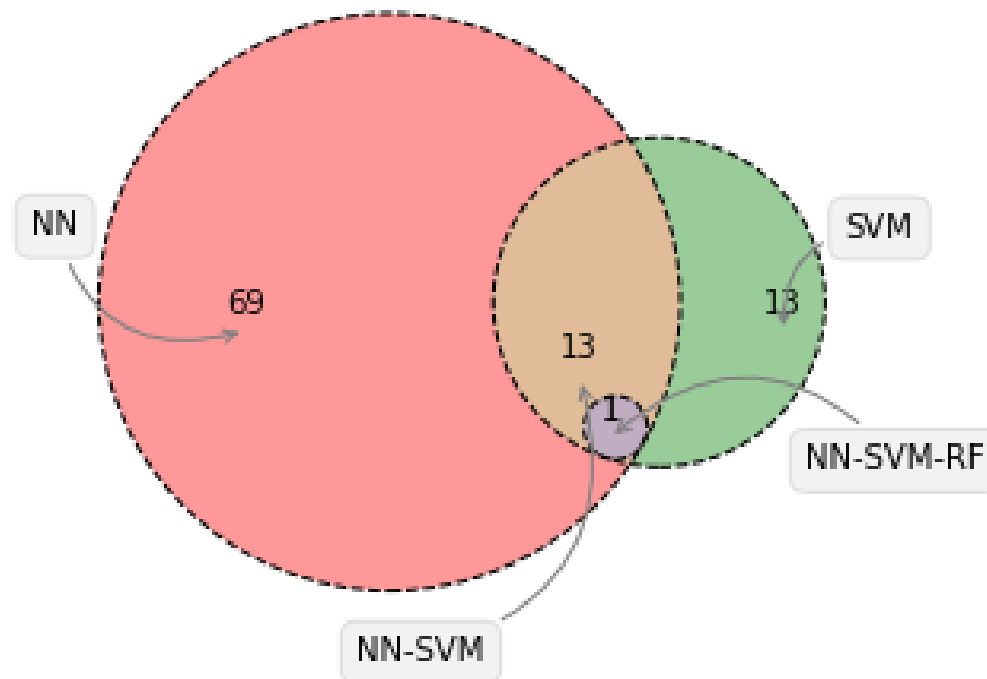
実測不適合 & 予測適合



- 3つの機械学習モデルのデータの83.3%が重複
- 誤った結果: **共通の要因**

実測適合 & 予測不適合

実測不適合 & 予測不適合



- 3つのモデルで同じデータ: 1件
- NNとSVMで同じデータ: 14件

- NNの正解件数が最も多く、かつ、SVMとRFの結果を**包含**している。
- 3つのモデルで同じデータ: 11件

3つ機械学習モデルの相関性・妥当性が確認できた。

- NN, SVM, RFを使用し, 実際の入力項目及びその検査の結果から予測モデルを構築し, その精度を評価した.
- 今回のデータ及び評価値ではNNが最も評価値の高い, 良いモデルであった.
- NNでは不適合の可能性の**高い上位2割の製品に不適合の約8割が存在した.**
- 上記の結果はあくまで今回のデータと評価値を使用した場合での結論であり, これらを変更した場合, 他の機械学習がより優れた結果を出す可能性がある.
- 3つの機械学習モデル間の関係を調査したところ, **良好な相関関係があることが確認できた.** 予測と実測が異なる場合においても各機械学習モデル間で同様の傾向が確認できた.

本研究は日本財団の助成のもと，日本船舶品質管理協会と協力して行われた．関係者各位に謝意を表します．

ご清聴ありがとうございます