

11 不適合納品物判定（品質管理）へのAIの適用

構造安全評価系 *馬 沖, 構造基盤技術系 平方 勝, 小沢 匠, 知識・データシステム系 間島 隆博

1. はじめに

多種多様な要素部品からなる工業製品の製造にあたっては、要素部品メーカーに発注することが多く、その要素部品が納品されるタイミングで検査を実施する。この検査では、大量の要素部品を取り扱うため、検査効率化の観点から、事前に不適合となる製品を予測する価値は高い。

現在、人工知能（Artificial Intelligence: AI）は第三次ブームを迎えており、ニューラルネットワークをはじめとする機械学習が世間から注目を集めている。造船海事産業においても、国土交通省が推進する海事生産性革命（i-Shipping）にてビッグデータやAIに関して言及されている¹⁾。機械学習は、事前に用意した学習データに対して統計処理等を施すことにより、データ間の関連性や規則を抽出し、自身のアルゴリズムを発展させる手法である。統計解析で行われるような分析にとどまらず、判断材料となる予測結果を提示すること、学習により予測精度が向上するなど、活用が期待される。

本研究では実際の検査の過程で得られた取引先や製品の型番などの入力情報及びその時の検査の適合/不適合の結果がまとめられた膨大な情報から機械学習により適合/不適合を予測するモデルを構築し、その精度に関して検証を行った。

2. 機械学習モデル

機械学習には囲碁や将棋などで活用され、話題になった強化学習、そして、データマイニングなどで活用されている教師なし学習などがある。本研究の機械学習では取引先や製品の型番などの入力情報の他に、実際に検査員が判断した適合/不適合の結果に基づいて予測モデルを構築するため、教師あり学習に該当する。また、結果は連続値ではなく適合/不適合のいずれかのため二項分類問題となる。教師あり学習においてもその学習方法はガウス過程や決定木など多岐にわたり、その精度はデータの非線形性などに依存する。そのため、実用上は複数の学習方法で精度評価を行い、精度が高いものを適用する必要がある。本研究では学習方法の他に後述する各学習方法におけるパラメータの最適化にも注力するため、学習方法は「ニューラルネットワーク（NN）²⁾」、「サポートベクターマシン（SVM）³⁾」及び「ランダムフォレスト（RF）⁴⁾」の3つを対象に、検討を行った。

2.1 学習方法

本節では本稿で使用する各学習方法の概要を記載する。なお、NN、SVM及びRFのアルゴリズムはPythonのライブラリ関数keras、sklearn.svm及びsklearn.ensembleをそれぞれ使用した。

(1) NN

本稿で使用する多層パーセプトロンタイプのNNモデルを図-1に示す。図-1中の丸はユニット、縦に表示したユニットの集合を層と呼ぶ。

図-1中の一番左の層を構成しているユニットは入力ユニットと呼び、本稿においては取引先や製品の型番などの入力情報を定量的に変換したものである。また、中央2つの層は隠れ層と呼ばれるものであり、この層の数が多いものがいわゆる深層学習（ディープラーニング）と呼ばれる。一番右は出力ユニットであり、本稿では適合/不適合をそれぞれ予測するユニットを設けた。最終的な予測値は出力ユニットでの結果を踏まえた0以上1以下の値となる。

2層目以降のユニットは一層上のユニットから重み関数などを踏まえた上で求められる。NNでは繰り返し計算により、教師データの誤差が最小化するように、ユニット間の重み関数を最適化する。

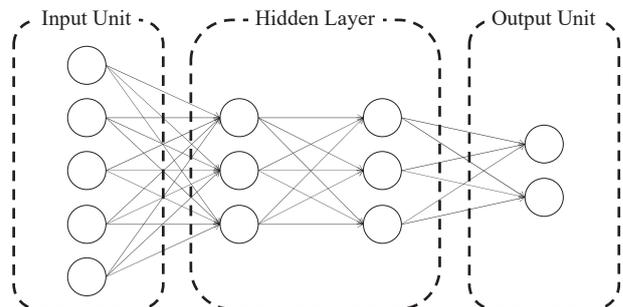


図-1 NNモデル構造図

(2) SVM

SVMは二項分類問題を解くために考えられた学習方法である。本稿で使用したSVMは非線形のものではあるが、説明を簡略化させるため、線形でのSVMのモデルを図-2に示す。図-2において、 x_i は入力情報を定量化したデータ集合であり、 w は各入力情報の重み関数である。図-2に示すように、SVMでは入力情報を定量化したデータ集合を変数とした識別関数（識別超平面）を作成し、この識別関数にて分けられた領域にて二項を分類する。ただし、本稿では不適合側の領域を正、適合側の領域を負とし、絶対値を識別関数からの距離とした連続値を予測結果として用いる。

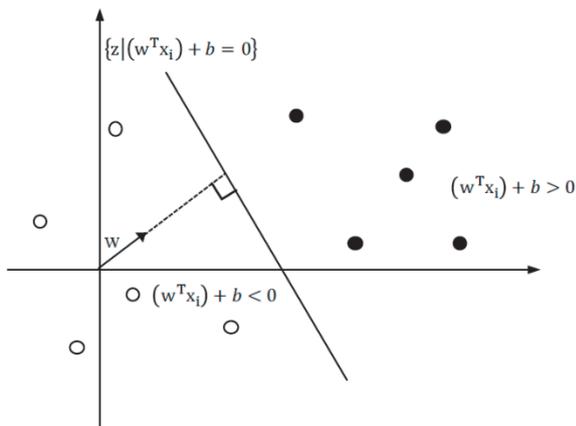


図-2 線形 SVM モデル構造図

(3) RF

RF は分類, 回帰, クラスタリングといった目的で使用される決定木を用いた学習方法である. 図-3 に教師データのサンプル及びそれを用いて作成した決定木を示す. 図-3 に示すように決定木では教師データから結果を予測する入力データの条件分岐を最適化し, モデルを構築する.

RF では教師データを復元抽出でランダムに選択して作成した複数の教師データから決定木をそれぞれ作成し, 各決定木の結果の平均値を RF の結果として使用する. なお, 本稿では 100 本の決定木を作成した.

本稿の目的は適合/不適合の分類問題ではあり, RF は前述のとおり分類問題にも対応できるものの, 本研究では回帰問題を扱う RF を使用することとした. 回帰問題では連続値を扱えるためより一般的であり, 不適合の度合いと言った数値的な出力が可能となる.

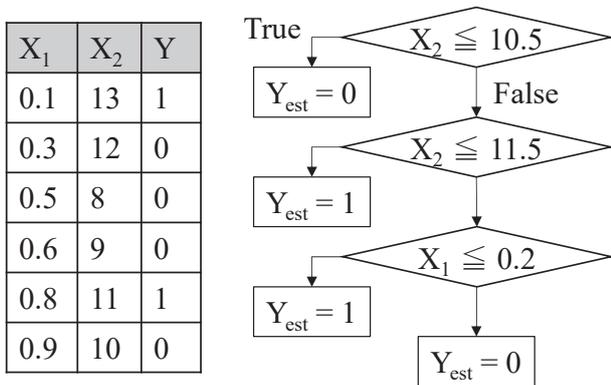


図-3 RF 計算例

2. 2 入力データの変換

NN や SVM では多項式の関数を使用するため, 入力データは全て数字である必要がある. 一方, 製品の情報は単価や注文数など定量的な値の他に取引先や装置コードなどの文字情報, 注文日や納期などの日付情報といった定量的な値で表現出来ない情報, 及び定量的な値であってもデータに欠損があるものもある. このようなデータは機械学習で取り扱えるよう, 定量的な値に変換する必要がある. 本稿で使用した実際の検査で得られたデータでもこのような入力情報が存在し

た. 本節では本研究にて用いたこれら定量的な値のみで表現できないデータの変換方法に関して記載する.

(1) 文字情報

文字情報は全体に対するそのデータの占める割合に着目し, 割合の高いデータの数だけ変数を新たに設ける. 新たに追加した変数はそれぞれ上位を占める各データに該当するか否かを 1/0 で表す. この変換例を表-1 に示す. 表-1 において A 社, B 社及び C 社が占める割合の高いデータ, Z 社が占める割合の低いデータとなる. 占める割合の高いデータは該当する変数のみ 1 であり, 他は 0 となる. また, 占める割合の低いデータは変数全てが 0 となる. なお, データに欠落がある場合, 欠落したのもも独立した文字として扱う. 欠落したデータの占める割合が高い場合, 他と同様に新たに追加する変数にて空白のデータか否かを 1/0 で表現する.

(2) 日付情報

注文日や納期などの日付情報は厳密には基準日からの日数を用いて定量的な値として取り扱うことが出来る. しかし, 将来的な予測においてはこの数値は今後, 教師データのものより未来の値で取り扱う, すなわち外挿することとなる. また, 注文日や納期などは年や日より月が重要となるケースが多いと考え, 本稿では月のみに着目した上で前項の文字情報と同様に扱った.

(3) 定量的な値

単価や注文数などの定量的な値かつデータの欠落がない場合, 入力情報の値をそのまま使用する. 一方, 定量的な値ではあるが, データの欠落がある場合, 数値をその値域内で数段階程度に区分し, (1) の文字情報と同様, その区分に該当するか否かで 1/0 で表現した. 変換例を表-2 に示す.

表-1 入力データ変換例(文字情報)

変換前	変換後		
	取引先: A 社	取引先: B 社	取引先: C 社
A 社	1	0	0
B 社	0	1	0
C 社	0	0	1
Z 社	0	0	0

表-2 入力データ変換例(データ欠落)

変換前	変換後		
	100 以上	100 以上 10 未満	10 未満
201	1	0	0
50	0	1	0
2	0	0	1
(欠落)	0	0	0

2. 3 学習方法の評価

NN では, 隠れ層の数や学習パラメータを最適な値に変更す

る等、SVM では、識別関数の非線形性を増す等、学習方法を複雑化させることにより、教師データにおける誤差を小さくする工夫をこらす。一方、学習方法の複雑化はその予測モデルが教師データに特化してしまうことなどから未知のデータに対する予測精度が落ちる、すなわち汎化性能が低下する。これは一般的に過学習と呼ばれる。過学習になっていないかをモニターするため、本稿では入力情報とその結果が与えられたデータの内、90%を教師データ、残りの10%を検証データに分けた。教師データで予測モデルを作成した後、その予測モデルを用いて検証データの入力情報からその結果を予測し、実際の検証データの結果と比較した。

この過学習を防ぐための各学習方法のパラメータ最適化及び各学習方法の比較において、予測モデルの精度を定量的に比較する必要がある。本稿では予測モデルの精度を定量的に表現した値を評価値とし、これを Precision と Recall の積と定義した。ここで、Precision とは不適合と予測されたもののうち、実際に不適合であったものの割合であり、Recall とは実際に不適合であったもののうち、不適合と予測されたものの割合である。この評価値の定義はあくまで本研究にて暫定的に取り決めたものである。重要度は各部品の発見率 (Recall) に高低を持たせることで対応できるものと考えられる。

予測モデルの結果が今回のように連続値である場合、適当な閾値を定めた上で、閾値との大小関係により適合／不適合を判定する必要がある。また、先ほどの評価値は、たとえ同じ予測モデルを使用したとしてもこの閾値の値に依存する。そのため、本稿では一つの予測モデルに対して評価値が最大となる閾値を求め、その閾値における評価値のみを使用する。

2. 4 入力データの重要度評価

前節で述べたとおり、適切な予測モデルの構築において、過学習の防止は重要である。この過学習を防止する手法の一つとして、重要度（寄与度）の低い入力情報を削除する方法が考えられる。重要度の低い入力情報の削除は過学習を防止するだけでなく、計算コストの削減にも繋がる。そこで、入力情報の重要度を評価する必要がある。RF では Python のライブラリ関数 (sklearn.ensemble) に重要度を評価するものがあるため、これを使用する。NN 及び SVM では次に示す手法で得られる決定係数を重要度の評価として用いる。

(i) 全ての入力情報を使用した訓練データを用いて予測モデルを構築し、検証データでの予測値を求める。

(ii) 検証データの入力情報において、項目ごとに入力情報を変更した検証データを新たに作成する。例えば表-1 の場合、注文数などはそのまま変えず、取引先を A 社、B 社、C 社及びその他 (Z 社など) に変えた全ての組み合わせの検証データを作成する。

(iii) 検証データ変更前後の予測値の関係における決定係数を求める。

ここで、決定係数が高いことはその入力項目が変更された

としても予測結果が大きく変わらない、すなわち重要度が低いことを示す。逆に決定係数が低いことは重要度が高いことを示す。なお、本来であれば NN と SVM のそれぞれで入力項目の重要度を評価すべきであるが、本稿では時間の都合上、NN で評価した重要度を SVM においても適用した。

3. 予測結果

3. 1 入力データ

本稿で使用する入力項目及びその種類と内容を表-3 に示す。本稿で使用したデータは 25 項目の入力データによって構成される。また、製品ごとのデータの総数は約 37 万件であり、うち 0.5%にあたる約 2000 件が不適合の製品であった。なお、RF においてのみ、代表者生年は予測に寄与しないと判断し、予め入力項目から削除した。

表-3 入力項目内容

入力項目	種類	内容
取引先	C	発注したメーカーのコード
機種	C	製品が使用されるものの種類
装置	C	製品が使用される機種の部分
細分の有無	V	製品コードでの細分の有無
素材の有無	V	製品コードでの素材の有無
注文日	D	発注した日付
納期	D	注文の納期の日付
注文数	V	注文した数量
取消数	V	注文数の内、取り消した数量
初回答納期	DB	メーカーが最初に回答した納期
回答納期	DB	メーカーが回答した最新の納期
単価	V	発注した製品の単価
発注金額	V	今回の注文で発生する金額
発注 LT	V	注文日から納期までの稼働日
部品 LT	V	製品の作成にかかる日数
検査区分	C	検査のあり/なし及びその区分
初回品マーク	V	初めての発注か否か
発注累積回数	V	メーカーがその製品に対し現在までに発注した回数の累計
発注累積個数	V	メーカーがその製品に対し現在までに発注した個数の累計
資本金	VB	メーカーの資本金
従業員数	VB	メーカーの従業員数
評点	VB	メーカーの評価点
代表者生年	VB	メーカー代表者の生年
不適合累積件数	V	メーカーがその製品に対し現在までに不具合を発生させた総数
不適合累積金額	V	メーカーがその製品に対し現在までに不具合を発生させた総額

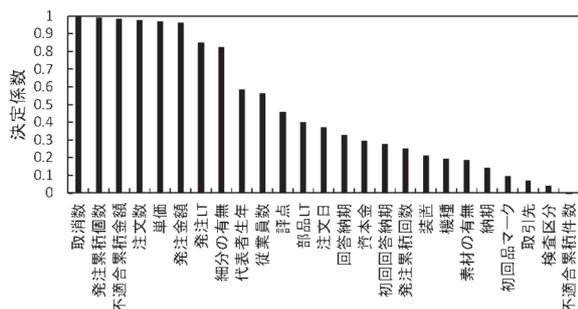
C: 文字データ D: 日付データ
V: 数量データ B: データの欠落あり

3. 2 入力項目の重要度

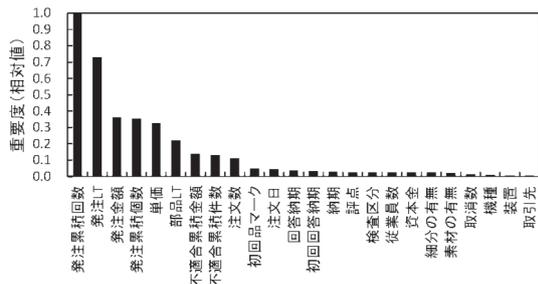
NN における各入力項目の決定係数及び RF における各入力項目の重要度を図-4 に示す。ここで、2.4 節で述べたとお

り、NNの重要度は決定係数の低い方が高いため、図-4(a)においては右側の入力項目の重要度が相対的に高い。したがって、NNの場合、不適合累積件数が最も重要度が高い結果となっている。一方、図-4(b)においては縦軸が重要度のため、左側の入力項目の重要度が相対的に高い。したがって、RFの場合、発注累積回数が最も重要度が高い結果となっている。

図-4の重要度を比較した場合、このように両者の傾向は必ずしも一致しない。これは2.4節で述べたとおり、NNとRFの重要度は異なる評価方法を用いていること、及び多項式関数近似に基づいたNNと条件分岐に基づいたRFは根本的に異なる予測手法であることが原因であると考える。



(a) NNに関する決定係数



(b) RFに関する重要度

図-4 入力項目に関する決定係数と重要度

この結果に基づき、各学習方法にて検討した使用する入力項目の組み合わせを表-4及び表-5に示す。NNとSVMの場合、表-4のように、図-4(a)における入力項目の決定係数順に区別した。全ての入力項目を使用する場合をCase 1とし、重要度の低い入力項目を一部削除した場合をCase 2とし、重要な入力項目のみを使用した場合をCase 3として検討した。RFの場合、表-5のように、図-4(b)における入力項目の重要度順に区別し、全ての入力項目を使用する場合をCase Iとし、重要度の高い入力項目のみを使用する場合をCase IIとし、重要度の低い入力項目のみを使用する場合をCase IIIとして検討した。

表-4 入力項目一覧 (NNとSVM)

入力項目	Case 1	Case 2	Case 3
取消数~代表者生年	●		
従業員数~初回答納期	●	●	
発注累積回数~不適合累積件数	●	●	●

表-5 入力項目一覧 (RF)

入力項目	Case I	Case II	Case III
発注累積回数~注文日	●	●	
回答納期~取引先	●		●

3.3 評価値

表-6に各学習方法の評価値及び最適パラメータを示す。ここで最適パラメータとは2.3節で記載した過学習を防ぐ最適なパラメータである。NNでは表-6の最適パラメータ欄に示した入力項目、隠れ層の数及び学習回数(重み関数を求めるための繰り返し計算の回数)であり、SVMでは表-6の最適パラメータ欄に示した入力項目及び識別関数を表現するパラメータC(誤分類をどの程度許容するかを決めるパラメータ)と γ (ガウシアンカーネルの直径の大きさ)の値である³⁾。RFでは表-5の入力項目がこの最適化させるパラメータに該当する。表-6の最適パラメータ欄に示したように、表-5に示したCASE Iの条件であった。本稿では、最も精度良く適合/不適合を判別出来た学習方法はNNであった。この最適パラメータにおける実測と予測の関係ごとの件数及び割合を表-7に示す。表-7に示すとおり、実際に不適合であったものを不適合と予測出来た割合はNNが最も高く、これがNNの高い評価値に貢献したと考える。一方、実際に不適合であったものを不適合と予測出来た割合が次いで高かった学習方法はSVMであるものの、評価値はSVMが最も低い。これはRFにおいて実際は適合であったものを誤って不適合と予測したものが1件のみであったこと、すなわち精度よく判定できたことが原因として考える。これによって評価値の構成要素であるPrecisionが、RFの場合、高くなったことが原因と考える。

表-6 各学習方法の評価値及び最適パラメータ

学習方法	評価値	閾値	最適パラメータ
NN	0.0803	0.1910	入力項目: Case 3 隠れ層: 4層 学習回数: 400
SVM	0.0538	-0.9995	入力項目: Case 3 C = 40 $\gamma = 0.0123$
RF	0.0688	0.8600	入力項目: Case I

表-7 実測と予測の関係ごとの件数及び割合

実測	不適合		適合	
	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%)

2.3節で述べたとおり、本研究はPrecisionとRecallの積を評価値として各学習方法の最適パラメータ及び閾値を求めた。参考まで、この最適解において閾値を変更した場合のRecallを図-5に示す。図-5において横軸は全件数のうち、不適合と予測したものの割合である。横軸の値が大きくなるにつれ、閾値が低くなり、不適合と予測されるものが増えるため、Recallの値が増加する。図-5（図-6は図-5の拡大図である）に示すとおり、横軸の値が10%、すなわち全件数のうち、不適合の可能性が高い上位10%を調べることにより、NN及びRFで約70%、SVMで約50%の不適合品が発見出来ると期待出来る。この値は一般的に検査業務の効率化に十分資すると考える。また、学習を重ねるにつれ、全ての学習方法がより精度を向上させると思われる。

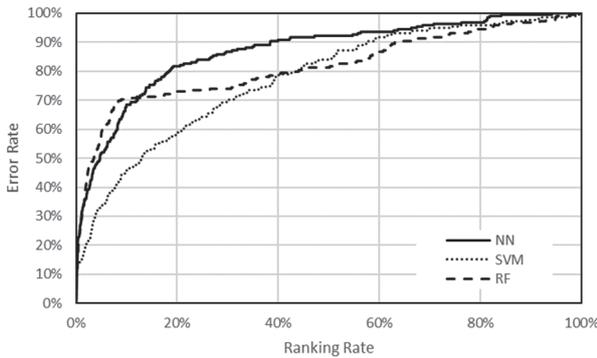


図-5 Ranking RateとError Rateの関係図

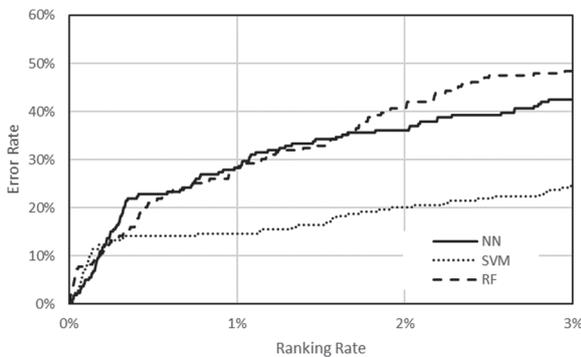


図-6 Ranking RateとError Rateの関係図（拡大図）

本稿で検討した3つの機械学習モデルは、全く異なる機械学習理論に基づいている。本稿では、同じ教師データを用いて、3つの機械学習モデルを訓練した。それぞれの機械学習の学習が満足いく結果であれば、同じ検証データを対象とした実測と予測結果について、3つの機械学習で大きな差が生じないものと思われる。そして、3つ機械学習モデルの相関性と予測結果の妥当性が確認できると思われる。

図-7から図-10では、表-7に示された4パターンについて、実測と予測結果の関係を集合関係であらわしたものであり、3つの機械学習モデルでの予測結果の関連性を示す。

図-7は、実測が適合であり、かつ、予測結果が適合であったパターンが、3つの機械学習モデルの結果の99.7%以上

で同じケース（図-7で薄紫色に表示）であったことを示している。すなわち、3つの機械学習は、このパターンにおいて、ほぼ同じ予測結果を出力している。

図-8は、実測が不適合であったにもかかわらず、適合と予測したパターンである。予測結果としては満足いくものではないが、3つの機械学習モデルの結果の83.3%で同じケース（図-8で薄紫色に表示）あったことを示している。高い確率であり、3つの機械学習モデルは満足いくものの、入力条件に限界があると考えられる。同じ入力条件においても、実測が適合であったり不適合であったりするようことが考えられ、何らかの特別な要因で、実測で偶然不適合であったことが考えられる。NNによって予測されるケースはSVMとRFの結果に包含されている。

図-9は、実測が適合であったにもかかわらず、予測結果が不適合であったパターンについて示す。予測結果としては満足いくものではない。3つの機械学習モデルで同じ結果を出力したケース（1件）を図-9で薄紫色に表示している。NNとSVMで同じ結果を出力したケース（13+1件）を橙色に表示している。

図-10は、実測が不適合であり、かつ、予測結果が不適合であったパターンについて示す。予測結果としては満足いく結果である。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。RFが最も保守的な結果になっている。

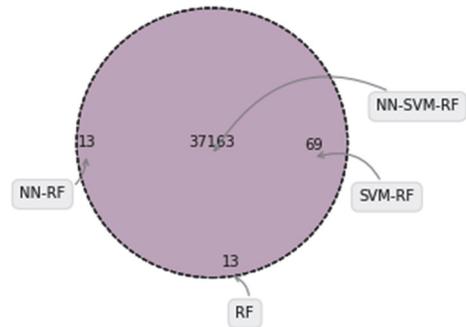


図-7 実測適合&予測適合

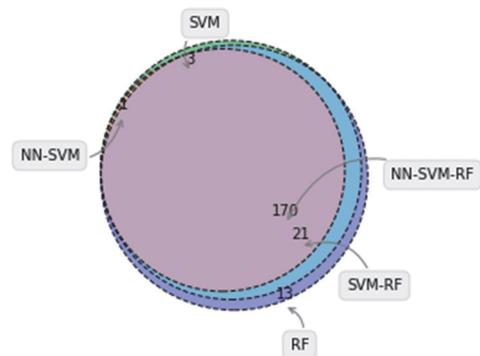


図-8 実測不適合&予測適合

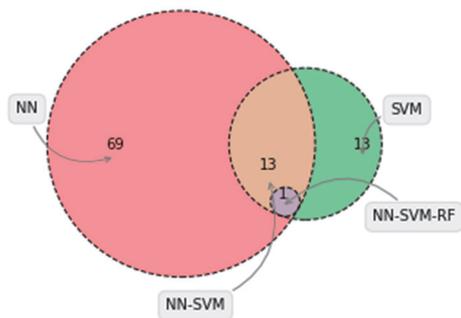


図-9 実測適合&予測不適合

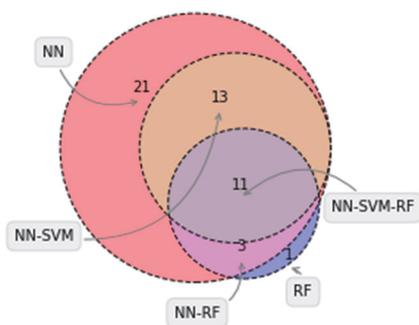


図-10 実測不適合&予測不適合

なお、上記の相関の議論は異なる閾値に基づいている。閾値が変更されると、予測結果の評価にも影響がある。そこで、閾値と関連なく、3つの機械学習モデル間の相関性をチェックするために、予測した不適合確率に従った3つのモデルの予測結果の並びを変えた上で、同じ Ranking Rate での予測結果の相関性を調査した。結果を表-8に示す。2つの機械学習モデルに相関性がない場合には、全データ N 件から M 件のサンプルを2回ピックアップすると、2回の選択中に同一のケースの件数が K を超える確率は、次のように解くことができる。

$$P(n > K) = \sum_{i=K}^M \frac{{}^M C_i \times {}^{N-M} C_{M-i}}{{}^N C_M}$$

上記の式によると、表-8に示す二つの条件 ($N=37478$, $M=94$, $K=21$ と $N=37478$, $M=375$, $K=52$) の発生確率は $1E-10$ よりはるかに小さい。したがって、3つの機械学習モデル間で、良好な相関関係があることがわかった。

図-8で考察したように、ある機械学習モデルで不適合と高い確率で予測するケースについては、他の2つの機械学習モデルにおいても、不適合と予測する確率が高くなると考える。

表-8 NN, SVM と RF の予測結果の相関性 (全ケース)

データ数	NN-SVM-RF	NN-SVM	NN-RF	SVM-RF
94 件 (0.25%)	21 件	41 件	25 件	36 件
375 件 (1%)	52 件	71 件	113 件	77 件

4. 結言

NN, SVM 及び RF を使用し、同じ入力情報を基に適合/不適合の学習、予測を行った。いずれの機械学習モデルについても、学習パラメータを最適化した。

本稿では、NN が最も精度の良い予測モデルとなった。ただし、この結果はあくまでこのデータ及び評価値を使用した場合における結論である。いずれの機械学習モデルも、教師データを蓄積し、さらなる学習を行っていくこと、さらに最適なパラメータを追求していくことで、予測精度は改善する余地がある。

それぞれの機械学習モデルの予測結果の考察に加えて、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。一方で、図-8のように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。これは学習に使用した入力データの不備などが原因として考えられるため、今後この信頼性に関して検証する必要がある。

今後は、社会実装を念頭に、これらモデルをベースにしたシステム開発を行っていく予定である。

謝辞

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。

参考文献

- 1) 国土交通省：海事生産性革命 (i-Shipping) の全体像
- 2) 岡谷貴之：深層学習，講談社，2015。
- 3) 小野田崇：サポートベクターマシン (知の科学)，オーム社，2007。
- 4) Breiman, L., "Random Forests", Machine Learning, Vol. 45, Issue 1, pp. 5-32, doi:10.1023/A:1010933404324, 2001.