

AI を利用した不良品の予測

小沢 匠*, 間島 隆博**, 平方 勝*, 馬 沖****

Prediction for Non-conformance Productions with Artificial Intelligence

by

OZAWA Takumi, MAJIMA Takahiro, HIRAKATA Masaru
and MA Chong

Abstract

We built machine-learning prediction models which predict pass/fail result of inspection on the basis of three hundred seventy thousand actual inspection results of parts for marine industrial product. This paper indicates the building process and accuracy evaluation of these models. We considered neural network (NN), support vector machine (SVM) and random forest (RF) as learning method, and optimize information used, constitution of hidden layer, number of learning and hyper parameters as well as optimization for prediction model building. We utilized learning data after converting to quantitative value, because of various learning data format, such as character string, value, date and these format with lack. As a result of information importance evaluation in the process of prediction models building, importance degree of non-conformance parts cumulative number, inspection classification and suppliers are high in NN, and that of order cumulative number, order lead time and order money amount are high in RF. Optimizing prediction model with product of precision and recall as evaluation value, accuracy of NN is highest and the top 10% of parts ,that are likely to be non-conformance, have 70% actual non-conformance parts. Furthermore, investigation of correlation between predicted values of each learning method shows that correlation between NN and SVM is high, but between RF and others are low. Therefore, we are of the view that RF predicts non-conformance parts on a different basis.

* 産業システム系, ** 知識・データシステム系, *** 構造安全評価系

原稿受付 令和 2年 1月 21日

審査日 令和 2年 2月 21日

目 次

1. まえがき	26
2. 機械学習モデル	26
2.1 学習方法	27
2.1.1 NN	27
2.1.2 SVM	27
2.1.3 RF	28
2.2 情報の変換	28
2.2.1 文字情報	29
2.2.2 日付情報	29
2.2.3 数量情報	29
2.3 予測モデルの評価	29
2.4 情報の重要度評価	30
3. 予測結果	30
3.1 学習に用いる情報	30
3.2 情報の重要度	30
3.3 評価値	33
3.4 学習モデルの予測値の相関	34
4. まとめ	34
謝辞	35
参考文献	35

1. まえがき

多種多様な部品からなる船用工業製品の製造にあたっては、部品メーカーに発注することが多く、その構成部品が納品されるタイミングで検査がしばしば行われる。この検査では、大量の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部品を予測する価値は高い。

一方、人工知能 (Artificial Intelligence: AI) は現在第三次ブームを迎えており、AI と密接な関りがある機械学習も世間から注目を集めている。また、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⁾。機械学習の手法の一つである教師あり学習は、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、例えば製品の異常の有無を予測し、それに伴う検査の必要性などの判断材料を提示することなど、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情報及びその時の検査の可否の結果がまとめられた膨大な情報を活用した教師あり機械学習により部品の適合／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る強化学習、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情報の他に、実際に検査員が判断した可否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、各学習方法のアルゴリズムは Python のライブラリ関数 (Sequential, svm.SVC, RandomForestRegressor) を使用した。

2.1.1 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1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。

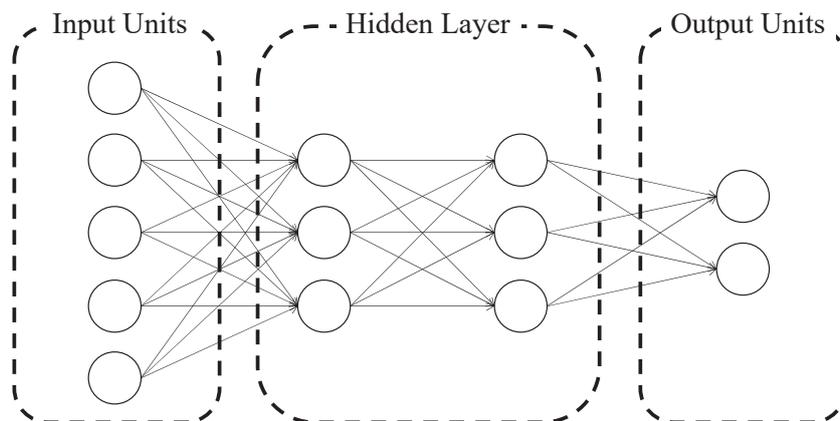


図1 NNモデル構造図

2.1.2 SVM

SVMは二項分類問題を解くために考案さ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 は情報を定量化したデータ集合であり、 w は各情報の重み関数である。図2に示すように、SVMでは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、この識別関数にて区分けさ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。

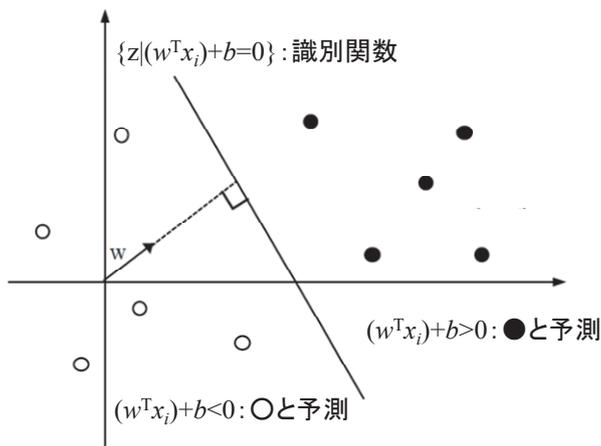


図2 線形 SVM モデル構造図

2.1.3 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. ここで, X_1 と X_2 はそれぞれ予測に用いる変数, Y はその変数の組み合わせにおける実績値, Y_{est} はその予測値である. 図3に示すように決定木では教師データから情報の条件分岐を最適化し, 結果を予測するモデルを構築する.

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あり, RF は先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量的な検討が可能となる.

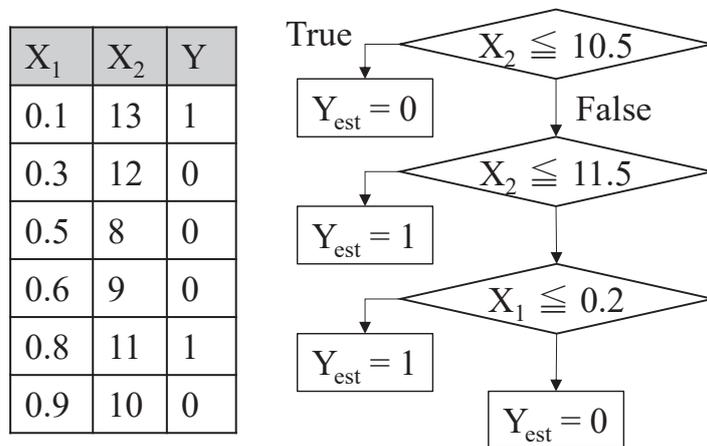


図3 決定木の構成概要

2.2 情報の変換

NN や SVM では多項式の関数を使用し, また RF においては情報の大小関係により決定木を構築するため, 使用する情報は全て数値として取り扱う必要がある. 一方, 部品の情報は単価や注文数など数値で直接表現出来る数量情報の他に, 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数値で直接表現出来ないものがある. 加えて, 例え数量的な情報であったとしても, データに欠損があるものも直接活用することはできない. そのため, このようなデータは機械学習で取り扱えるよう, 数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情報が存在した. 本節では本研究にて用いた数値で直接表現できない情報の変換方法に関して記載する.

2.2.1 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、他と同様に新たに加える変数にて空白のデータか否かを1/0で表現する。

表1 文字情報変換例

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

2.2.2 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できる。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り、精度の低下が懸念される。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。

2.2.3 数量情報

単価や注文数などの数量情報かつデータの欠落がない場合、この値をそのまま使用する。一方、数量情報ではあるが、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、2.2.1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2 欠落を有する数量情報の変換例

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

2.3 予測モデルの評価

NNでは隠れ層の数や学習パラメータを増やす、SVMでは識別関数の非線形性を増やす等、予測モデルを複雑化させることにより、一般に教師データにおける誤差は少なくなる。一方、予測モデル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。過学習を防ぐため、本稿では学習データの内、90%を教師データ、残りの10%を検証データに無作為に分けた。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。場合によっては「発見率 (Recall) 90%での労力 (不適合と予測される件数) を最小化する」などといった別の評価値での運用も可能である。

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 情報の重要度評価

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) の低い情報を削除する方法が考えられる。重要度の低い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一方、各情報の重要度を事前に評価する必要がある。RF では Python のライブラリ関数に重要度を評価するもの (`feature importances`) 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) 検証データの各情報において、項目ごとに値を変更した検証データを新たに作成する。例えば表 1 の取引先の重要度を評価する場合、注文数など他の情報はそのまま変えず、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間の決定係数を求める。

ここで、決定係数が高いことはその情報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その情報が変更された場合、予測結果が大きく変わる、すなわち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る方がより良いが、本稿では NN と SVM はともに多項式に基づく予測手法であることを踏まえ、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 学習に用いる情報

本稿で使用する学習データに記載されている情報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 項目の情報によって構成される。また、学習データの総数は約 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、事前の検討により代表者生年は予測に寄与しないと判断し、予め学習に用いる情報から削除した。

3.2 情報の重要度

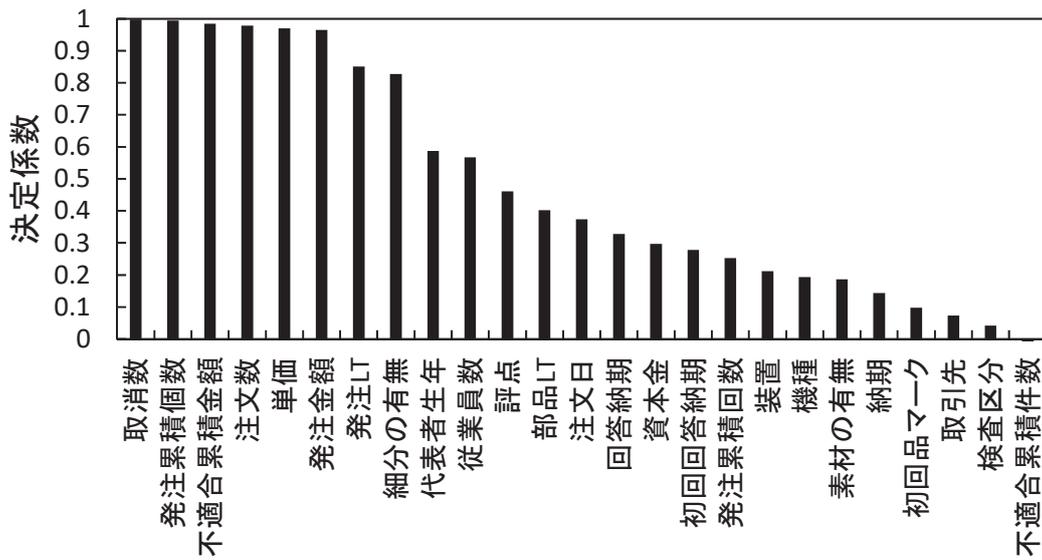
NN における各情報の決定係数及び RF における各情報の重要度を図 4 に示す。ここで、2.4 節で述べたとおり、NN の重要度は決定係数の低い方が高いため、図 4(a)においては右側の情報の重要度が相対的に高い。したがって、NN の場合、不適合累積件数が最も重要度が高い情報である。一方、図 4(b)においては縦軸が重要度のため、左側の情報の重要度が相対的に高い。したがって、RF の場合、発注累積回数が最も重要度が高い情報である。

この結果に基づき、各学習方法にて検討した使用する情報の組み合わせを表 4 及び表 5 に示す。NN と SVM の場合、表 4 のように、図 4(a)における各情報の決定係数順におよそ等分割な 3 つに区分けした。全ての入力項目を使用する場合を Case 1 とし、重要度の低い情報を一部削除した場合を Case 2 とし、重要な情報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)における各情報の重要度順におよそ半々に区分けし、全ての入力項目を使用する場合を Case I とし、重要度の高い情報のみを使用する場合を Case II とし、重要度の低い情報のみを使用する場合を Case III として検討した。

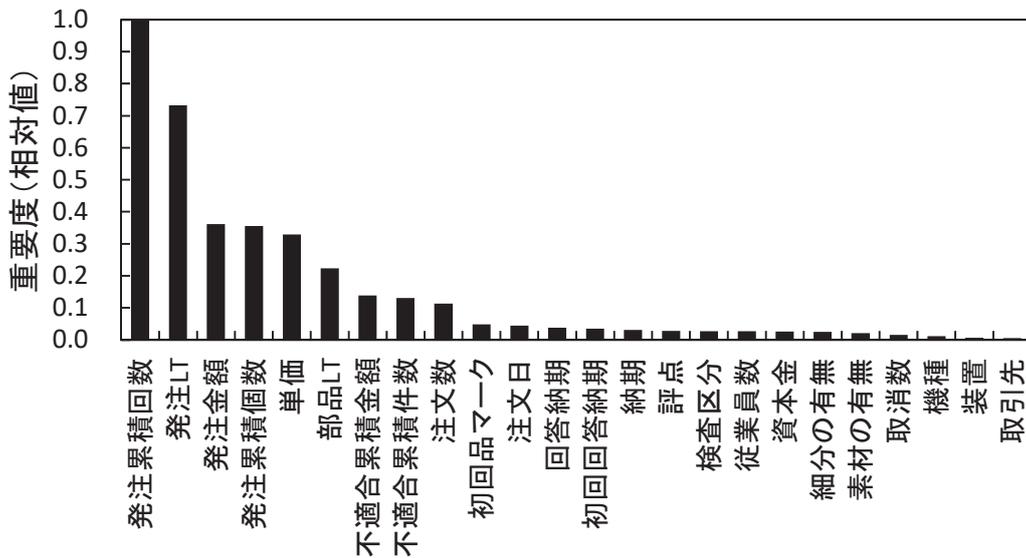
表3 学習データに記載されている情報の種類及びその内容

情報	種類	内容
取引先	C	発注したメーカーのコード
機種	C	部品が使用されるものの種類
装置	C	部品が使用される機種の部分
細分の有無	V	部品コードでの細分の有無
素材の有無	V	部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	発注した部品の単価
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	部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	メーカーがその部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	メーカーがその部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	メーカーがその部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	メーカーがその部品に対し現在までに不具合を発生させた総額

C：文字情報；D：日付情報；V：数量情報；B：データの欠落あり



(a) NNにおける決定係数



(b) RFにおける重要度

図4 各情報の重要度評価の結果

表4 NNとSVMにおける使用する情報の組み合わせ

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

表5 RFにおける使用する情報の組み合わせ

情報	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節に記載した過学習を防ぎ、検証データにて評価値が最大となる各学習モデルにおける設計変数の値である。NNでは使用する情報、隠れ層の構成及び学習回数（重み関数を求めるための繰り返し計算の回数）、SVMでは使用する情報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 （識別関数の複雑さを決めるガウシアンカーネルの直径の大きさ）の値³⁾、及びRFでは使用する情報がこの最適化させるパラメータに該当する。

表6に示すとおり、本稿の問題で最も高い精度を示した学習方法はNNであった。各学習方法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件のみである、すなわちPrecisionが高い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	使用する情報：Case 3 隠れ層：4層 学習回数：400
SVM	0.0538	-0.9995	使用する情報：Case 3 C = 40, $\gamma = 0.0123$
RF	0.0688	0.8600	使用する情報：Case I

表7 各学習方法における合否の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
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%)

表6及び表7は評価値が最大となる閾値における結果である。一方、閾値を高くした場合、不適合と予測される件数が増えるとともに、実際に不適合であるものの発見率（Recall）も上がる。この関係も学習方法の評価において重要であるため、閾値を変化させた場合の上記関係を図5に示す。

図5のとおり、NNやRFでは不適合の可能性の高い上位10%に実際に不適合である部品が70%含まれることになる。一般にこの予測精度は検査の効率向上に十分資すると考える。

また、NNとRFを比較した場合、横軸が10%以下の場合RFの発見率が高く、10%以降ではNNが高くなる。このことは最適な学習方法は学習データの内容の他に評価値の定義によって変化することを示す。そのため、表6に示したとおり、今回使用した評価値ではNNが最も高いという結果になったものの、この結果は評価値などで変化することに留意する必要がある。

なお、10%以降のRFのグラフが直線となっているが、これはRFが検証データの90%にて0と予測し、データが飛んでいることが原因である。

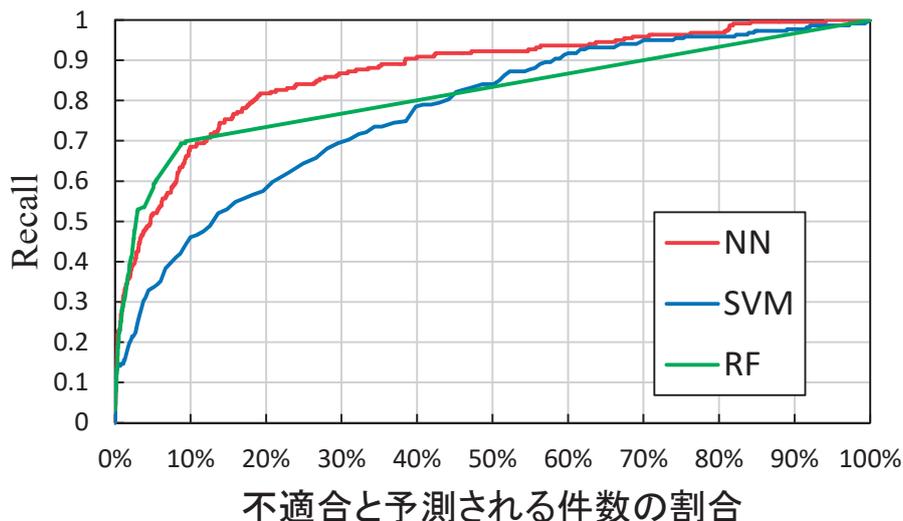


図5 閾値を変化させた際の不適合発見率の変化

3.4 学習モデルの予測値の相関

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、先述のとおり同じ教師データを用いて、3つの機械学習モデルを構築した。一方、それぞれの予測結果に相関があれば異なる予測モデルであったとしても同じ根拠に基づき不適合を予測したと考えられる。そこで、NN、SVM及びRFの各予測モデルにて計算された予測値の順位を計算し、各予測モデル間の順位の相関係数を求め、これを比較した。

結果を表8に示す。一般に相関係数が0.2~0.4であれば弱い相関、0.4~0.7であれば相関、0.7以上であれば強い相関があると考えられている。NNとSVMの相関係数は0.66であり、相関があることから同じ根拠に基づき不適合を予測したと考えられる。一方、RFに対するNNやSVMの相関係数はいずれも0.4未満であり、弱い相関であった。このことから、NNとSVMとの関係に比べるとRFは他の学習モデルと異なる根拠に基づき不適合を予測したと考えられる。これは2.1節に示したとおり、NNとSVMが多項式に基づく予測モデルであることに対し、RFは条件分岐に基づく予測モデルであることが原因と考える。

表8 各学習モデルの予測値における相関係数

学習モデル	相関係数
NN-SVM	0.66
SVM-RF	0.24
RF-NN	0.33

4. まとめ

NN、SVM及びRFを使用し、同じ情報に基づき検査の可否を予測する機械学習モデルを構築した。

本稿では、NNが最も精度の良い予測モデルとなり、不適合の可能性の高い上位10%に70%の不適合品が存在した。この予測精度は検査の効率向上に十分資すると考える。

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の相関関係を調査したところ、NNとSVMの相関に対してRFの相関が低い結果となった。これはNNとSVMが多項式に基づく予測モデルであることに対し、RFは条件分岐に基づく予測モデルであることが原因と考える。

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.