

数値化した画像データの機械学習手法による解析

～SVM とランダムフォレストの適用～

(株) 日本科学技術研修所
数理事業部
山田 芳幸

1. はじめに

近年、計算機の処理速度の向上や通信技術の発達があり、これまでは長い時間を必要とした高次元データの解析や大容量データの高速転送などが可能となった。そのため、画像の解析に関する技術が著しく発達した。中でも、畳み込みニューラルネットワーク(CNN)を中心としたDeepLearningと呼ばれる手法に関する解析手法はモデル構築の自由度の高さや予測性能の高さもあり、広く活用が見込まれている。

しかし、DeepLearning に関する手法の多くはモデル構築の自由度が高いために最適なモデルの設計が困難であることや、日々より効率的なモデルのパターンが報告されていることもあり、十分に使いこなすことは困難である。また、DeepLearning に関する手法は多くの場合、非常に多くのハイパーパラメータと呼ばれるモデル構築にあたり事前に最適な値を指定する必要のあるパラメータの最適化をする必要があるため、

- 計算に非常に時間が必要であり、高価な GPU を必要とすることが多い
- ハイパーパラメータが多く、過学習（既知のデータにのみ正確に予測でき、未知のデータの予測が上手く行えない状態）しやすい。対策として、非常に多くのデータを必要とする

といった欠点が存在する。他にも、「予測性能の向上」を主な目的とする機械学習手法が共通して持つこととなる「人が理解できない」という欠点も他手法に比べて強くなりやすい。

上記のような欠点があるため、画像解析によく用いられている DeepLearning に関する多くの手法は品質管理分野への適用が難しいとされているが、画像は非常に多くの情報を持つことから画像の解析は品質管理分野においても有用であると考えられる。

そこで、DeepLearning に関連しない機械学習手法を活用して画像の解析を試みた事例の報告を行う。

報告を行う事例は青山学院大学 理工学部 小野田 崇 教授を主査とした品質管理学会公募研究会「製造業のためのビッグデータの解析あり方研究会」の活動で使用された事例となり、解析に使用した画像や各種ヒアリング内容、目視点検による正解率などの事例に関する全ての情報は研究会で得た情報となる。

品質管理学会公募研究会「製造業のためのビッグデータの解析あり方研究会」は機械学習や AI の専門家を交えて現場の技術者が機械学習やビッグデータを利活用するためにはどのようにしたらよいかを実際のデータ解析も踏まえて議論を行う研究会であり、2017年10月から2019年9月の2年間にわたり活動していた。

2. 事例概要

2.1. 概要と目的

報告を行う事例は検査員が目視点検により、装置の傷み具合を3段階で分類していた「ある装置の外観検査の画一化」を試みた事例となる。

検査員が目視点検ということもあり、基準を定めたとしても人の「見る」という感覚を正確に数値化し、共通なものとして表現することが難しい。そのため、検査員によって傷み具合の診断基準に差異が生じることが懸念されていた。

他の問題点としては、1つの装置に対して熟練した検査員が複数人で判断すると約95%で同一の判断をするが、経験の浅い検査員だと約80%の割合でしか熟練した検査員と同一の判断を行えないという熟練度によっても正確さに差異が生じることがあげられていた。こちらは、熟練した検査員の養成を行うことで解決はできるが、検査員の養成や技術の継承には時間がかかってしまうために安易に解決できる問題ではなかった。

そこで、熟練した検査員の感覚を反映した形で熟練した検査員と同等の正確さで誰でも判断を行えるようにするため機械学習を用いてデータ解析・予測モデルの構築を行うことを目的とする。

本事例での解析のフローとしては図1に示すような手順で解析を行い、予測モデルを作成した。

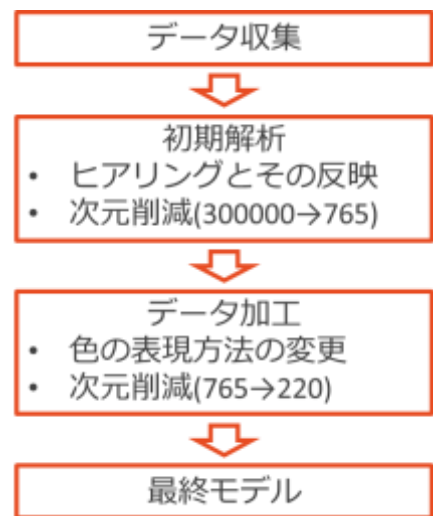


図 1. 解析のフロー

2.2. データ概要

人の「見る」という感覚を置き換えるため、解析には装置の目視点検箇所を撮影した画像を対象とする。撮影した画像は320pixel×320pixelの大きさになり、撮影した画像の一部を図2に示す。

解析対象は画像ではあるが、画像を直接解析すること一般的には困難となる。そこで、画像を座標毎にR, G, Bの3色、それぞれ0から255までの整数に変換したものを解析することとした。そのため、解析を行うデータの次元は320×320×3と非常に大きな次元となった。

また、撮影された装置の傷み具合は熟練した検査員が画像を確認し、3段階(level1, level2, level3)



図 2. 解析に使用した画像の一例

に分類してある。

データ数は、最も傷み具合の軽度となる level1 と判断された装置の画像が約 750 枚、軽度の傷みがあると判断された level2 が約 400 枚、最も傷んでいると判断された level3 が約 450 枚の計 1600 枚となる。

3. 初期解析

3.1. 熟練した検査員へのヒアリング

まず初期解析を行う前に、熟練した検査員へのヒアリングを実施した。ヒアリングの目的としては画像からではわからない検査員の知見を確認するためである。そのために、ヒアリングを行った検査員には解析のために撮影した画像を見てもらい、傷み具合の判断も行ってもらった。また、判断の際にどのように画像を見て、傷み具合を判断したのかを重点的に答えてもらった。

熟練した検査員へのヒアリング結果からは、

- 傷み具合に関する部分のみを見ている
- 位置関係や形状は見ておらず、画像全体で判断している
- 明るさは特に気にしていない

と大きく3つの点で判断していることが分かった。

ヒアリングした結果から位置関係や形状に関する観点から座標情報がいらなことが明らかであるが、解析に使うために数値化した画像データは座標毎にデータを持っているため、位置情報を含んでしまっている。そこで、次元の削減を行うために解析に使うデータから座標情報を削除することとした。

座標情報を削除し、画像全体の情報とするデータの編集は色の情報のみのデータに編集することと同じとなる。今回は色の情報のみにデータを編集する方法として、R・G・Bの各色の値の出現回数を新しく変数とすることとした。その結果、次元数を $320 \times 320 \times 3$ から 256×3 と大幅な次元数の削減となった。

3.2. 新しいデータの確認

次元を大幅に減らすことができたため、可視化による確認を行うことが容易となった。そこで、傷み具合毎に各色の値の出現回数の算術平均を算出し、確認を行った。可視化した結果が図3となる。

図3を確認したところ、

- 傷み具合が level1 の時は3色とも非常に正規分布に近い形で色が出現している
- 傷み具合がひどくなるにつれて R,G,B の差異が大きくなり、分布の形状が正規分布から離れていく

といった傾向があることが分かる。このことから、熟練した検査員へのヒアリングした通り、画像から位置情報を抜いて色の情報だけに変換したデータでも機械的に分類できそうであることが推測できるため、今後の解析には色の情報だけを用いていくこととした。

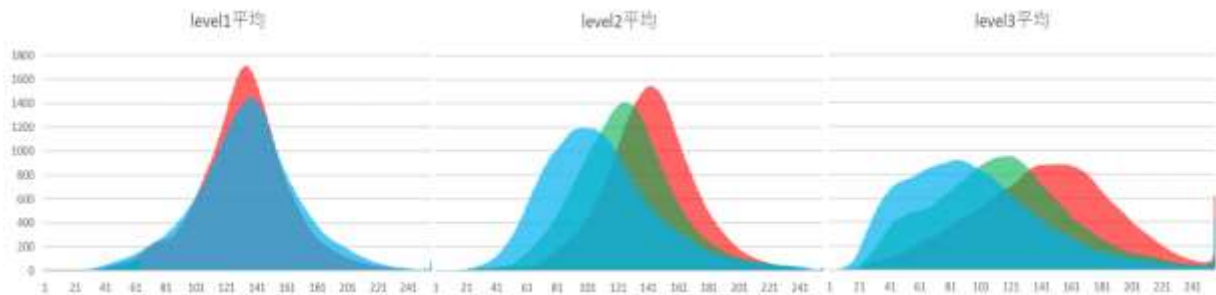


図 3. 傷み具合毎の色の平均出現回数 (図中の色は R,G,B に対応)

3.3. RGB 色空間でのモデルの作成

図 3 を確認し、色の情報だけで 3 段階の傷み具合を分類できそうであると推測できた。しかし、色に関する全ての情報の中には分類に不要な情報があることも考えられたため、分類に不要な情報を削除したいが、誤って分類に必要な情報を削除してしまうと予測の正確さが低下すると考えられる。

そこで、分類に必要な情報を誤って削除したことがわかるように、色に関する全ての情報がある状態で作成した予測モデルの予測性能を評価し、予測性能のベンチマークを作成することとした。ベンチマークの作成には全ての色の情報を使ったモデルの作成を行う。

今回、予測モデルの作成に使った機械学習手法はカーネル法を用いて高次元空間を作成し、作成した空間で線形判別を行うサポートベクターマシン (以下, SVM) と、決定木 (AID に非常に類似した手法) を多数作成し、その後にそれらの多数決で予測を行うランダムフォレストの 2 手法となる。また、ハイパーパラメータの評価とモデルの評価は共に 5 分割交差検証法を行った。

SVM を用いて作成した予測モデルの誤判別表を表 1 に示す。表 1 の誤判別表を確認すると学習データ (予測モデルの作成に使用したデータ) の正答率は約 98.8% と非常に高く、傷み具合の判断基準をしっかりと理解できていると思われる。

既存のデータから判断基準を明確に理解することも重要だが、新しく検査した時の正確さの方も重要となる。そこで、新しく検査した時の正確さを評価するためのテストデータ (予測モデルの作成に使用していないデータ) の正答率を確認すると約 78.8% と低く、新しいサンプルに対しては新人と同程度の正確さでしか傷み具合を判断できないということが分かる。

表 1. SVM で作成した予測モデルの誤判別表

		正答			合計	誤分類数	誤分類率(%)	適合率(%)	
		level1	level2	level3					
学習データ	予測結果	level1	743	6	0	755	6	0.795	99.205
		level2	7	425	1	433	8	1.848	98.152
		level3	5	0	452	457	5	1.094	98.906
		合計	761	431	453	1645			
	誤分類数	12	6	1		19			
	誤分類率(%)	1.577	1.392	0.221			1.155		
	再現率(%)	98.423	98.608	99.779				98.845	
テストデータ	予測結果	level1	656	119	42	817	161	19.706	80.294
		level2	85	262	32	379	117	30.871	69.129
		level3	20	50	379	449	70	15.590	84.410
		合計	761	431	453	1645			
	誤分類数	105	169	74		348			
	誤分類率(%)	13.798	39.211	16.336			21.155		
	再現率(%)	86.202	60.789	83.664				78.845	

表 2. ランダムフォレストで作成した予測モデルの誤判別表

		正答			合計	誤分類数	誤分類率(%)	適合率(%)	
		level1	level2	level3					
学習データ	予測結果	level1	713	15	0	728	15	2.000	97.940
		level2	41	407	26	474	67	14.135	85.865
		level3	7	9	427	443	16	3.612	96.388
		合計	761	431	453	1645			
	誤分類数	48	24	26		98			
	誤分類率(%)	6.307	5.568	5.740			5.957		
再現率(%)	93.693	94.432	94.260				94.043		
テストデータ	予測結果	level1	642	79	15	736	94	12.772	87.228
		level2	92	304	57	453	149	32.892	67.108
		level3	27	48	381	456	75	16.447	83.553
		合計	761	431	453	1645			
	誤分類数	119	127	72		318			
	誤分類率(%)	15.637	29.466	15.894			19.331		
再現率(%)	84.363	70.534	84.106				80.669		

次に、ランダムフォレストを用いて作成した予測モデルの誤判別表を表 2 に示す。ランダムフォレストを用いた予測モデルでも SVM を用いた予測モデルの時を予測性能の傾向は同じであることが分かる。

これら 2 手法の予測モデルの誤判別表から、色に関する全ての情報を用いて予測モデルを作成した場合、予測の正確さは新人程度の約 80%にしかならないということが分かり、作成した予測モデルの傷み具合の分類基準は熟練した検査員の分類基準と異なると思われる。このことから、熟練した検査員が無視している不要な情報を予測モデルの中に取り込むことで誤った分類基準を作成していることに起因すると考えられ、予測の正確さを熟練した検査員と同等にするためには不要な情報の洗い出しと削除を行うこととする。

4. 情報の取捨選択

分類に不要な情報の洗い出しのために再度、熟練した検査員へのヒアリング結果について考察を行った。その結果、熟練した検査員へのヒアリング結果の中でも対策を打ちやすい位置情報に対してしか対策を打っていないことが明らかであった。そこで、まだ取り入れていない熟練した検査員へのヒアリング結果を反映させることによって不要な情報の削減を目指した。

4.1. 色空間の変更

初めに行ったのは色空間の変換である。色の表現方法には様々な方法があり、画像データは一般的に RGB 色空間という方法で情報を保存している。しかし、人間の感覚だと画像は R・G・B の 3 色の色で構成されるのではなく、色合いや色の鮮やかさ、明るさといった情報として取り入れている。そして、今回は熟練した検査員の感覚を反映した予測モデルの作成が目的である。これらのから、分析に使う色空間は RGB 色空間ではなく、より人の感覚に近い HSV 色空間 (図 4 にイメージを示す) にすることとした。

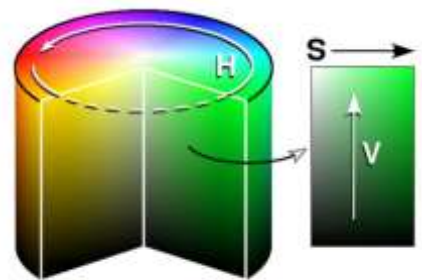


図 4. HSV 色空間のイメージ

HSV 色空間は RGB 色空間から非線形な変換によって得られ、画像を色相 (H) と鮮やかさ (S)、明度 (V) で表現し、色相は 0 から 360 までの値をとり、 $H = 0$ の時と $H = 360$ の時で同じ色を持つような値であり、鮮やかさは 0 から 100 の値をとり、値が大きいほど鮮やかな色であることを表し、明度は 0 から 100 の値をとり、値が大きいほど明るい色

であることを表す。

HSV 色空間にデータを変換することで明るさを表す値を得ることができた。ここで、熟練した検査員へのヒアリング結果を確認すると「明るさは特に気にしていない」と答えているので V については不要な情報として解析に使わないこととした。

4.2. 色相の指定

熟練した検査員は傷み具合の判断時には傷み具合に関する部分のみを見ているとヒアリングでは答えていた。そこで改めて数値ではなく画像そのものを確認していると、傷み具合が進行するにつれて赤系列の色が強くなっていることに気が付いた。そこで HSV 色空間の色合いを示す H に対して解析で使う領域を制限することとした。色相を確認したところ、赤色から黄色 ($0 \leq H < 60$) と紫色から赤色 ($300 \leq H < 360$) の領域に赤色が含まれていると判断できたため、 $0 \leq H < 60, 300 \leq H < 360$ の範囲に指定して傷み具合毎にヒストグラムを作成した (図 5)。

図 5 を確認したところ、傷み具合毎に傾向がことになっており、予想通り傷み具合が進むにつれて赤色に近づいていることが分かる。このことから、熟練した検査員も同じような色を中心に見ているのではないかと判断し、色相については $0 \leq H < 60, 300 \leq H < 360$ の範囲を解析に使うこととした。

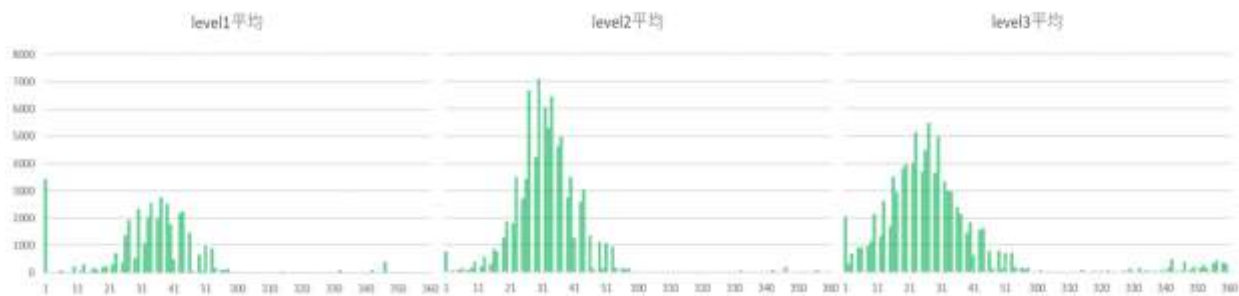


図 5. 赤色を含むと思われる色相のヒストグラム (傷み具合別)

4.3. 鮮やかさの指定

鮮やかさについては熟練した検査員へのヒアリングから解析に使うかの判断が難しい。傷み具合毎平均ヒストグラムを作成し傾向の確認を行い、分類に使うかを判断した。作成したヒストグラムは図 6 になる。

図 6 を確認したところ、傷み具合が進行すると大きな値の出現回数が増えることが分かる。鮮やかさは値が小さいほど色としては灰色に近くなるためこちらも画像の傾向と比較すると一致して

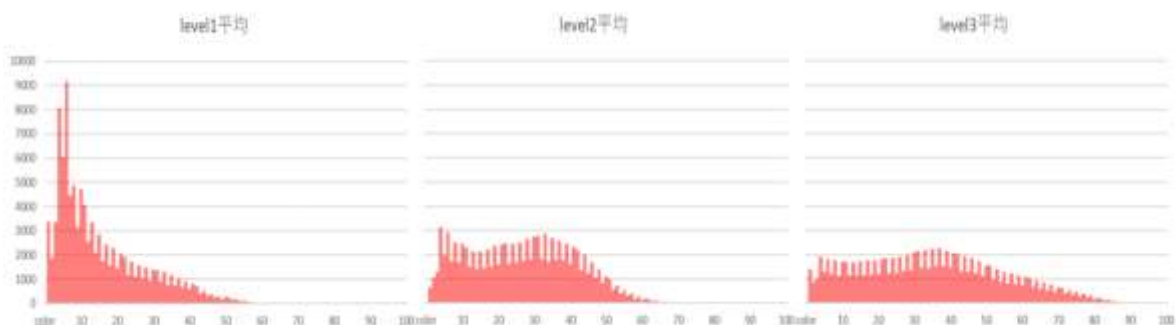


図 6. 傷み具合毎の鮮やかさのヒストグラム

いると考えられた。そのため、鮮やかさについては全ての値を使い、解析をすることとした。

5. 予測モデルの作成

情報の取捨選択での議論によって解析に使うデータの次元は256×3から 220 まで減らすことができた。この220次元のデータを用いてSVMとランダムフォレストによる予測モデルを作成した。

さらに、新しい予測モデルの作成を行う際には、データがわずかながら不均衡であることを考慮し、各傷み具合のデータ比率を保持するようにテストデータの抽出時に層化抽出を行った。

5.1. SVM による予測モデル

色の情報全てを使って予測モデルを作るときと異なり、ハイパーパラメータの評価を厳密に行った。そのためにはハイパーパラメータと誤分類率の関係をプロットした図 7 のグラフを確認した。グラフが3つあるが、SVM は本来 2 群の判別しか行えない手法であるのを今回は「level1 とその他」、「level2 とその他」、「level3 とその他」という形で3群の分類を行っているためである。

図7を見ると、3つのグラフとも極小値を取っているとは言えないものの値の変化が起きなくなっている部分から選択されている。このことからこれ以上大きなハイパーパラメータを探索しても予測性能の向上は見込めないと判断し、探索を終了した。

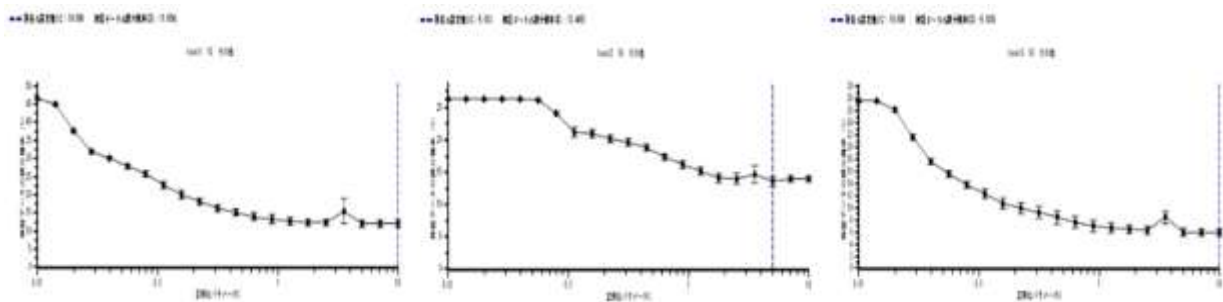


図 7. ハイパーパラメータと誤分類率の関係

ハイパーパラメータの評価を終えたので、得られた予測モデルの予測性能の評価を行うために誤

表 3. HSV 色空間で情報を削減した状態での予測モデル (SVM)

		正答			合計	誤分類数	誤分類率(%)	適合率(%)	
		level1	level2	level3					
学習データ	予測結果	level1	759	7	0	766	7	0.914	99.086
		level2	2	423	15	440	17	3.864	96.136
		level3	0	1	438	439	1	0.228	99.772
		合計	761	431	453	1645			
	誤分類数	2	8	15		25			
	誤分類率(%)	0.263	1.856	3.311			1.520		
	再現率(%)	99.737	98.144	96.689				98.490	
テストデータ	予測結果	level1	676	80	10	766	90	11.749	88.251
		level2	80	331	38	449	118	26.281	73.719
		level3	5	20	405	430	25	5.814	94.186
		合計	761	431	453	1645			
	誤分類数	85	100	48		233			
	誤分類率(%)	11.170	23.202	10.596			14.164		
	再現率(%)	88.830	76.798	89.404				85.836	

判別表を使い評価した。この時使用した誤判別表が表 3 になる。

新しく作成した予測モデルでは学習データの正答率が約 98.5%と色に関する全ての情報を使って得られた予測モデルと大きな変化はないが、テストデータの正答率は約 85.8%と大きく向上して

いることが分かる。

5.2. ランダムフォレストによる予測モデル

ランダムフォレストではハイパーパラメータは大きいほど予測性能が高くなりやすいものと、経験則的にある程度めぼしい値を見つけることのできるものしかない。そのため、ハイパーパラメータの調整を特に行わず誤判別表を用いて評価した。新しく作成した予測モデルの誤判別表を表4に示す。

ランダムフォレストでも SVM の時と同様に学習データの正答率に大きな変化はないが、テストデータの正答率には大きな向上が見られる。

表 4. HSV 色空間で情報を削減した状態での予測モデル (ランダムフォレスト)

		正答			合計	誤分類数	誤分類率(%)	適合率(%)	
		level1	level2	level3					
学習データ	予測結果	level1	734	9	0	743	9	1.211	98.789
		level2	26	414	18	458	44	9.607	90.393
		level3	1	8	435	444	9	2.027	97.973
		合計	761	431	453	1645			
	誤分類数	27	17	18		62			
	誤分類率(%)	3.548	3.944	3.974			3.769		
	再現率(%)	96.452	96.056	96.026				96.231	
テストデータ	予測結果	level1	667	61	2	730	63	8.630	91.370
		level2	93	332	47	472	140	29.661	70.339
		level3	1	38	404	443	39	8.804	91.196
		合計	761	431	453	1645			
	誤分類数	94	99	49		242			
	誤分類率(%)	12.352	22.970	10.817			14.711		
	再現率(%)	87.648	77.030	89.183				85.289	

5.3. 新しく作成した予測モデルの予測性能について

2つの手法を用いて新しく予測モデルの作成を行ったが、両者ともに色に関する全ての情報を使い作成したモデルに対してテストデータの正答率が大きく向上する結果となった。これは、反映していなかった熟練した検査員へのヒアリング結果を上手く反映した結果だと考えられる。

6. まとめ

機械学習を用いた画像の分類は画像を数値データに変換することで DeepLearning を使わずに行うことが可能であった。しかし、画像を数値化してしまうと座標毎の3つの要素を持つ非常に大きな次元のデータを解析する必要性が生じてしまった。また、大きな次元を持つが目的には不要な情報が混じることで予測性能が低下するとも考えられた。

そこで、今回の報告では熟練した検査員の感覚を反映させるという目的のために実施した熟練した検査員へのヒアリング結果をもとにして、情報の削減を行った。情報の削減を行った後のデータを使い予測モデルを作成したところ、十分に情報の削減を行えなかった時点で作成した予測モデルに比べて予測性能が約80%から約85%と約5%の大きな向上させることができた。

また、熟練した検査員へのヒアリング結果をもとに情報を削減したことにより、本来であればどのような基準で判断をしているかが分からなくなる機械学習に対して、人の感覚を反映させることができたと考えられる。

付録 本要旨中で使用する機械学習分野での用語について

本文中やシンポジウム当日での発表にあたり、機械学習分野で一般的に使われる単語をいくつか使用させていただいた。特にシンポジウム当日での発表では機械学習分野で一般的に使われる単語の解説を行いだすと非常に時間がかかることとなることもあり、省略させていただいた。

そこで本章にはそれらの単語の意味するところを簡単にまとめておく。

- ハイパーパラメータ
機械学習手法によりモデルを構築するために事前に決定しておくべきパラメータ。データから推測できる値ではないため、予測性能などを指標として最適な値を探索して決定する。
- 5分割交差検証法
k分割交差検証法と呼ばれる評価方法で分割数kを5としたもの。
分割したデータの1つを取り除いた残りのデータを使い、モデルを構築し、取り除いておいたデータの予測を行うことでモデルの予測性能を計算する。この作業を様々なハイパーパラメータの組み合わせに対して行い、取り除いておいたデータの予測が最も正確に行える組み合わせを採用する。
- 学習データ
k分割交差検証法でモデルの構築に使ったデータの総称。
- テストデータ
k分割交差検証法でモデルの予測性能の計算に使ったデータの総称。
- 汎化能力
未知のデータの予測が正しく行えるかを表す。機械学習ではこの能力をできる限り向上させることが目的となり、学習データの予測性能とテストデータの予測性能を使って評価する。
- 過学習
汎化能力が低く、学習データの予測しか行えないような状態。

7. 謝辞

研究会主査をしてくださった青山学院大学 理工学部 小野田 崇 教授を始めとする研究会メンバーの皆様にはお忙しい中研究会に参加し、非常に興味深い議論を交わしていただいたことに感謝します。

8. 参考文献

- RGB – Wikipedia , <https://ja.wikipedia.org/wiki/RGB>
- HSV 色空間 – Wikipedia
<https://ja.wikipedia.org/wiki/HSV%E8%89%B2%E7%A9%BA%E9%96%93>
- 色相 – Wikipedia , <https://ja.wikipedia.org/wiki/%E8%89%B2%E7%9B%B8>

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ください <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>