

# 全天雲画像の分類における撮影データ追加の効果

江越 航 \*

## 概要

科学館屋上に設置したライブカメラが撮影した全天雲画像を、機械学習を利用して分類する試みを継続して行っている。前報では、全天雲画像に加えて、解析雨量データや気象台の観測データを追加することで、より適切な分類が可能となった。一方、SHAP 値による分析からは、太陽の影響が大きいことが明らかになった。そこで、カメラの撮影データを追加することで、さらに適切な分類が可能か試みた。本稿では、その手法、および結果について述べる。

### 1. はじめに

近年、機械学習を用いた人工知能の技術が、飛躍的に進歩している。2024年のノーベル物理学賞は、機械学習の基礎を築いた業績に対して、そして化学賞はAIを活用したタンパク質設計に対して、いずれもAI分野の研究者が受賞している。[1]

機械学習とは、コンピューターが画像などの大量のデータからルールや知識を自ら学習する技術のことである。一方、天気予報においては、物理法則に基づく数値計算による予測が現在でも主流である。しかし近年、大手IT企業により、完全にAIだけで予測するシステムが発表されている。また、欧州中期予報センター(ECMWF)も機械学習による予測のための、学習済みのデータセットを提供し始めており、今後機械学習による予測が主流となる可能性がある。[2]

一方、気象台においては、目視による雲量観測を2024年3月で終了している。これは観測技術の向上に伴い、必要な情報が自動で得られるようになったためである。しかし、観測は天気予報の基礎となるものであるから、画像を解析して適切な情報を付与することができれば、気象の普及活動においてより有用と考えられる。

科学館では屋上にライブカメラを設置し、2014年5月から2021年11月まで、およそ7年半の全天雲画像のデータを取得した。しかしこの取得データの活用はされていない。そこでこの全天雲画像を、機械学習を利

用して分類する試みを継続して行っている。前報[3][4]に引き続き、画像の分類性能を向上させるための試みについて述べる。

### 2. 研究の概要

本研究は、科学館でこれまで蓄積してきた7年半に渡る全天画像・気象観測データを用いて、機械学習を用いて画像を解析し、分類を試みたものである。

前報においては、全天雲画像を入力データとし、ニューラルネットワークモデルを用いて機械学習により雲画像の分類を行ったところ、入力データとして解析雨量データや気象台のアメダス観測値を追加することで、分類の精度が向上した。さらに、太陽方位角・高度などの天文データを追加することで、より適切な分類が可能となった。

一方で、やはり学習用データに対する過学習が起り、未知の画像に対しては、十分な精度を得ることができないという問題が生じた。

また、SHAP 値による分析を行ったところ、昼間の時間帯の分類では太陽の影響が大きいことが明らかになった。

そこで今回、カメラ付属の計器により取得されていた上空温度や照度などの撮影データを利用することで、さらに適切な分類が可能か試みた。

以下の項目では、その手順と解析結果を述べる。

### 3. 撮影データ

図1は今回解析に用いている全天雲画像である。今までの報告では、この全天雲画像のみ、もしくは解

\*大阪市立科学館 学芸員  
e-mail:egoshi@sci-museum.jp

析雨量のデータや、近隣の気象台のアメダス観測値を合わせて解析に用いていた。しかし、この雲画像を取得しているカメラには、カメラ以外にもいくつかのセンサーが搭載されており、画像中には取得された気温や気圧などの数値が記載されている。

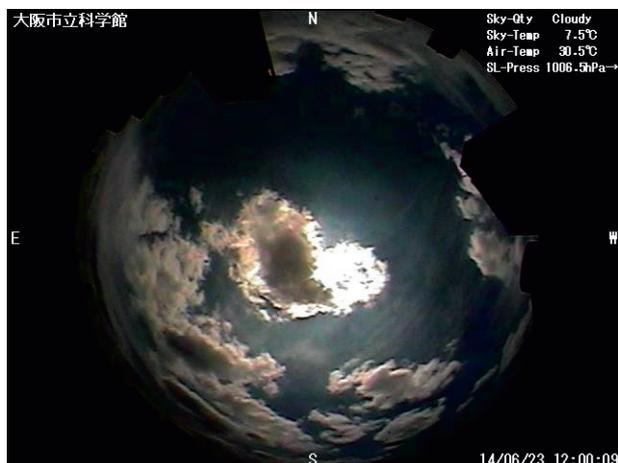


図1 全天画像

図2はカメラ付属のソフトウェアによって描かれた気象データである。外気温と上空温度、大気圧の変化が描かれている。いずれもカメラに付属しているセンサーが取得したデータをグラフにしたものである。

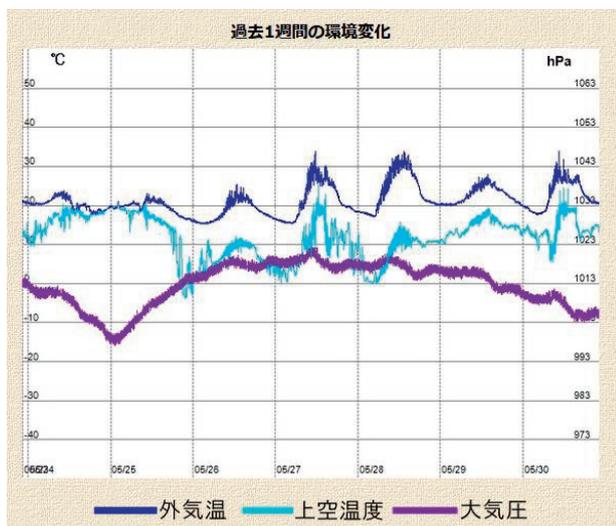


図2 観測データ

外気温は、熱電対で測定した温度であり、直射日光が当たることがある。そのため、通常の気温より高い値を示す。

また上空温度は、機器に上空に向けて取り付けられた赤外放射温度計が取得した温度である。これは、地上から対流圏上層部までの、平均の温度を示すと考えられる。快晴の時はささぎる物がないので、上空温度の表示温度が低くなる。逆に曇りの時は、雲の底の温度が計測される。つまり、上空温度と外気温の差は、

どのくらい曇りが出ているかの目安になる。そのため上空温度の値は、全天雲画像の分類の際に、有用な指標になることが予想される。

前報では、全天雲画像に加えて、解析雨量データや気象台の観測データ、太陽方位角・高度などの天文データを追加することで、より適切な分類が可能となった。その際、日照時間や太陽高度、全天日射量など、太陽が大きく寄与していることが判明した。

そこで本研究では、雲画像と合わせて、次項に示すカメラの撮影データも合わせて解析することを試みた。

## 4. データの準備

### 4-1. 訓練データ

入力となる訓練データは、前報でも用いた科学館で取得している図1の全天画像データである。元の画像サイズ 640×480 ピクセルの RGB カラーデータを、48×48 ピクセルに縮小している。

さらに、カメラの撮影データを合わせて入力することとする。カメラの撮影データとして記録されている項目は表1に示す通りである。

表1 撮影データの内容

項目		数値例
日付	Date	2014/5/23
時刻	Time	12
稼働状況	status	1
電源制御	ON_OFFs	200
絞り	irismode	1
外光	light	2
カメラ制御	MTVsetNO	2
ドーム温度	DomeTemp	34.1
内部温度	InsideTemp	33.5
外気温	OutsideTemp	23.9
電源電圧	SupplyVoltage	15.7
上空温度	SkyTemp	12.7
大気圧	Pressure	1014.1
照度	Illuminance	1470.388

ただし、この撮影データは、すべての期間において保存されていなかった。そこであらかじめ、全天雲画像が存在する日時と、カメラの撮影データが存在する日時を比較して、両者とも存在するデータを抜き出して、解析するようにした。

またさらに、前報でも用いた気象台の観測データ、および太陽と月の方位角・高度などの天文データを加えた。訓練に用いたデータの項目を、表2、表3に示す。

表2 訓練に用いたデータの項目(気象)

気圧(現地)	気圧(海面)	降水量	気温	露点湿度
蒸気圧	湿度	風速	風向	日照時間
全天日射量	降雪	積雪	視程	

表3 訓練に用いたデータの項目(天文)

太陽	月			
方位角	高度	方位角	高度	位相

入力データとしては、カメラの特性を考慮し、昼間の時間帯で、かつ教師データとなる雲量・天気観測データがある [6時、9時、12時、15時、18時] を用いることとした。

期間は比較的データ量の多い2015年の1年間、あるいは2014年～2021年の全期間のデータを用いて解析した。全天雲画像とカメラの撮影データの両方が存在するデータ数は、2015年の1年間の場合1673、2014年～2021年の全期間の場合は10,328であった。

#### 4-2. 教師データ

教師データとしては、先に述べたように、気象庁ホームページに公開されている、過去の気象データの大坂の観測値のデータから、「天気」および「雲量」のデータを用いた。天気については、[快晴、晴れ、薄曇、曇、雨、みぞれ、雪、雷電]の8種類、雲量については、0～10の11種類への分類を試みた。

### 5. 機械学習の実行

機械学習のモデルとしては、前報と同じく、入力データとして全天雲画像に加えて、別の入力データを合わせて用いることから、モデルを作成する際に複数の入力可能なFunctional APIを用いて作成した。

表4 ニューラルネットワークモデルの構成

層	パラメータ	出力サイズ	
		雲画像	撮影データ
入力層	画像データを1列に	6,912	14 or 33
中間層	ニューロン数 1,024	1,024	1,024
ドロップアウト	25%	1,024	1,024
結合		2,048	
中間層	ニューロン数 64	64	
出力層	ニューロン数 11 or 8	11 または 8	

機械学習の手法としては、表4に示すニューラルネットワークモデルを用いて、全天雲画像と撮影データ(14項目)またはさらに追加して気象台の観測データ、天文データを合わせた入力データ(33項目)から、天気または雲量の分類を試みた。なお、以下の解析結果では、雲量を分類した結果について述べる。

実際に学習を実行する際には、データ全体の9割を学習用データとし、1割はテストデータとすることで、過学習が起こっているかを確認するようにした。

### 6. 解析結果

#### 6-1. 雲画像と撮影データによる学習

最初に、雲画像と撮影データのみを組み合わせ、ニューラルネットワークモデルを使って機械学習により雲画像を分類して雲量の分類を試みた。

2015年のデータのみを用いて、学習回数(エポック数)100回で実行したところ、正答率は学習用データ68%、テストデータ46%であった。ただし、次項の図4のoriginalの分布から分かるように、教師データの雲量は0または10が多くを占めている。実際の分類結果を見たところ、特に雲量が1～9の中間の範囲ではあまりうまく分類することができていなかった。

そこで、使用するデータを2014年～2021年の全期間とし、さらに学習回数を500回に増やして実行したところ、正答率は学習用データ67%、テストデータ59%となった。十分に正答率が上昇したとは言えないが、前報のように、学習回数を増やすことで過学習となることはなかった。

#### 6-2. 雲画像と撮影データ・気象台の観測データによる学習

次に、雲画像と撮影データに加えて、気象台の観測データ、天文データを合わせて、ニューラルネットワークモデルを使って機械学習により雲画像分類して雲量の分類を試みた。

2015年のデータのみを用いて、学習回数(エポック数)100回で実行した場合、正答率は学習用データ70%、テストデータ43%で、前項の結果と大きく違わないが、雲量が1～9の中間の範囲の分類成果がよくなった。

さらに使用するデータを2014年～2021年の全期間とし、学習回数を500回に増やして実行した結果を、図3、図4に示す。この場合の正答率は学習用データ78%、テストデータ56%となっている。正答率が上昇するとともに、過学習となる傾向はかなり抑えられている。

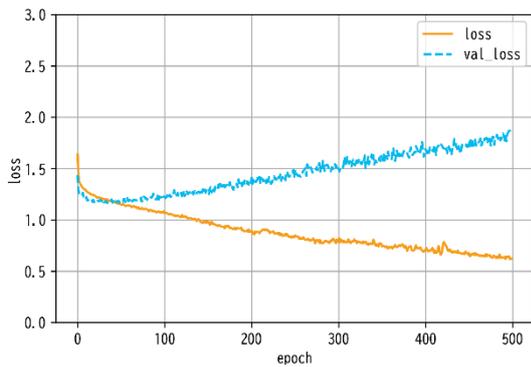


図3 損失関数の変化(撮影データ・観測データ追加)

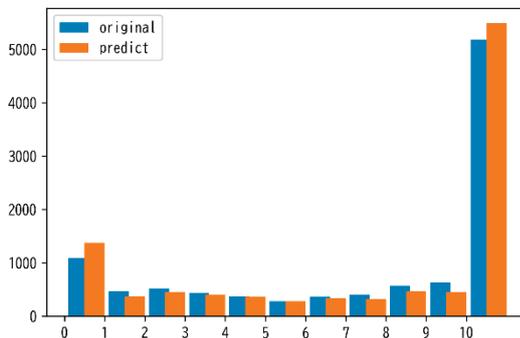


図4 雲量の分類(撮影データ・観測データ追加)

以上より、雲画像に加えて、撮影データと気象台の観測データ、天文データを追加して、特徴量を組み合わせて機械学習を行うと、ニューラルネットワークモデルであっても、より適切な分類が可能であることが分かった。特に撮影データを組み合わせることで、過学習を抑えながら、雲画像の分類が可能となった。

## 7. SHAP 値による分析

今回の解析においてどの気象要素が重要だったかを調べるため、SHAP 値による分析を行った。SHAP (Shapley Additive exPlanations)とは、予測値に対してそれぞれの特徴変数がどのような影響を与えたのかを算出する手法である

図5は雲画像と撮影データ、気象台の観測データ、天文データを組み合わせて学習させた際に、雲量を5と分類した場合の、各パラメータの寄与を解析したものである。特に上空温度(SkyTemp)の寄与が大きかったことが分かる。さらには、全天日射量や日照時間、照度(Illuminance)や太陽高度など、太陽に関する指標や、露点温度という雲発生に関わる指標の関連が深いことが分かる。

上空温度と照度は、カメラの撮影データから得られた値であり、雲画像の分類には、カメラの撮影データの値が有効であったことが分かる。

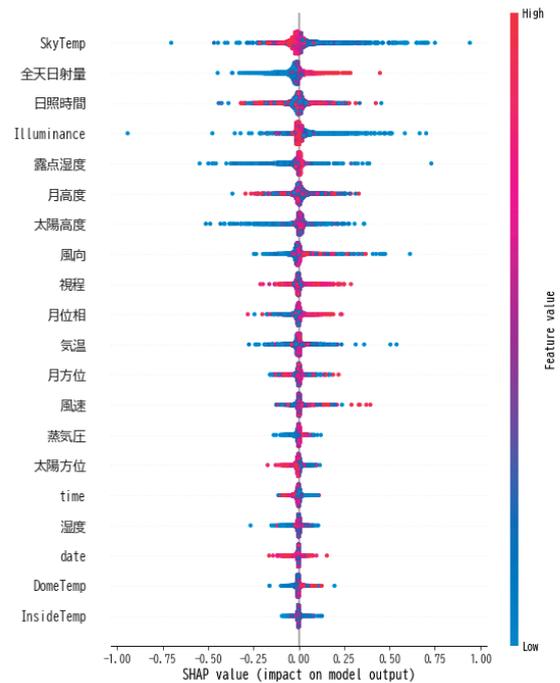


図5 SHAP 値による分析(雲量 5)

## 8. まとめ

科学館で蓄積してきた全天画像データを用いて機械学習により画像を解析し、雲量の分類を試みた。

入力データとして、雲画像に加えて、カメラの撮影データ、気象台の観測データ、天文データを合わせて、機械学習により雲画像から雲量の分類を行ったところ、ニューラルネットワークモデルであっても、適切な分類が可能となった。特に今回、カメラの撮影データを追加することで、学習用データに対する過学習が抑えられることが分かった。

SHAP 値による分析を行ったところでも、カメラの撮影データの影響が大きかったことが明らかになった。特に上空温度と照度の寄与が大きかったことから、これらの値の観測を行うことが、雲画像の分類にも有効であることが分かった。

## 謝辞

本研究はJSPS 科研費 JP22K02976 の助成を受けたものです。

## 参考文献

- [1] ノーベル賞公式ホームページ  
<https://www.nobelprize.org/all-nobel-prizes-2024/>
- [2] MOOC Machine Learning in Weather & Climate  
<https://learning.ecmwf.int/pages/index.html>
- [3]江越 航, 大阪市立科学館研究報告, 33, p1-4 (2023)
- [4]江越 航, 大阪市立科学館研究報告, 34, p5-10 (2024)