

解析雨量・アメダス観測値と全天雲画像の組み合わせデータへの 機械学習の適用

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。単に雲画像データのみで機械学習を行うと、学習用データに対する過学習が起こることから、未知の画像に対して十分な精度を得ることができない。そこで新たな気象データを追加して、特徴量を組み合わせて機械学習を行うことを試みた。特に今回は、周辺の気象状況も含む解析雨量のデータ、および近隣の気象台のアメダス観測データを用いて、全天雲画像データと組み合わせた解析を行った。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、飛躍的に進歩してきた。画像認識などの識別系 AI のみならず、近年は新しい文章や画像を作成する生成 AI も広く使われるようになってきている。

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。気象分野においても古くから用いられており、数値予報のデータを、ユーザーが利用しやすい形式に加工して、予報要素へ変換（翻訳）する、ガイダンスと呼ばれるものを作成する過程において、ニューラルネットワーク、カルマンフィルタ、ロジスティクス回帰などの機械学習を用いた手法が利用されている[1]。

筆者は前報[2]において、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、雲量や天気を分類する試みを行った。科学館の屋上に設置したライブカメラは、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ている。ただ単に画像を表示するだけでなく、画像を解析して適切に分類した情報を付与することができれば、より有用な内容になると考えられる。そこでこれに全天雲画像に機械学習を用いた解析を行うことを試みた。

しかし単に雲画像データのみで機械学習を行うのでは、未知の画像に対して十分な精度を得ることができなかった。そこで今回は、気象データを追加して、特徴量を組み合わせた機械学習を行うことを試みた。本稿では、その手法、および結果について述べる。

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、近隣の気象台で観測された雲量、および天気の観測結果を教師データとして、機械学習のためのニューラルネットワークモデル、および畳み込みニューラルネットワークモデル (CNN) を構築して、雲画像の分類を実行した。

しかし、単純に雲画像のみを入力データとした機械学習を行った場合、学習用データに対する過学習が起こり、テストデータに対する正答率が下がってしまう。そのため、未知の画像に対しては、必ずしも適切でない結果が得られる恐れがある。

そこで、雲画像と合わせて、別の気象データを入力データに加えることを試みることにした。今回、追加の入力データとして用いたのは、科学館周辺の降雨状況を示す解析雨量のデータ、および近隣の気象台のアメダス観測値である。

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

今回の解析では、全天雲画像に加えて、別の入力データを合わせて用いたことから、機械学習モデルを作成する際にも複数の入力可能なモデルを用いて作成した。

以下の項目では、その手順と解析結果を述べる。

3. 解析雨量ファイル

3-1. 解析雨量データの読み込み

分析に用いた解析雨量データは、気象業務支援センターよりオフラインデータを購入したものを用いた。このデータは、GRIB2形式のファイルで提供されており、およそ 1km メッシュ分解能で、毎正時 00 分および 30 分の前 1 時間降水量分布のデータとなっている。機械学習での分析に用いるためには、解析雨量のデータ形式を理解し、適切にプログラミングして必要範囲のデータを読み出す必要がある[3]。

解析雨量のメッシュデータ範囲は、東経 118 度～150 度、北緯 20 度～48 度で、経度 1 度を 80 等分、緯度 40 分(=1/1.5 度)を 80 等分した、南北 3360、東西 2560 メッシュデータから成る(表 1)。

ただし、この範囲のすべての地点のデータがある訳ではなく、実際には日本列島に沿った図 1 の範囲のみ、雨量データが存在する。それ以外のポイントは、-1 の値が入れられている。

表 1 解析雨量データの仕様

	経度	緯度
範囲	東経 118 度～150 度 (32 度分)	北緯 20 度～48 度 (28 度分)
メッシュ数	2560	3360
分割	経度 1 度を 80 等分 ($32 \times 80 = 2560$)	緯度 40 分を 80 等分 ($28 \times 60/40 \times 80 = 3360$)
1 メッシュ の角度	45 秒	30 秒
1 メッシュ の大きさ	約 1125m (25m/秒で計算)	約 900m (30m/秒で計算)

解析雨量のデータは、米国海洋大気庁(NOAA)が提供している wgrib2 と呼ばれるプログラムを用いて読み出すことができる。この際、グリッドポイントの番号は、北緯 20 度、東経 118 度を左下とするグリッドが 1 番目になる。

そのため、緯度、経度が与えられた時の、グリッドのインデックスの値は、以下の式で求められる。

$$\begin{aligned} \text{経度のインデックス} &: (\text{経度} - 118) / (1/80) + 1 \\ \text{緯度のインデックス} &: (\text{緯度} - 20) / ((1/80) \times (2/3)) + 1 \end{aligned}$$

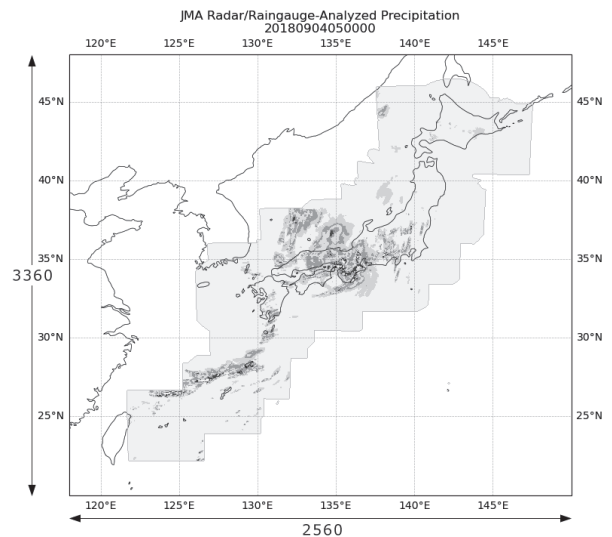


図 1 解析雨量データ

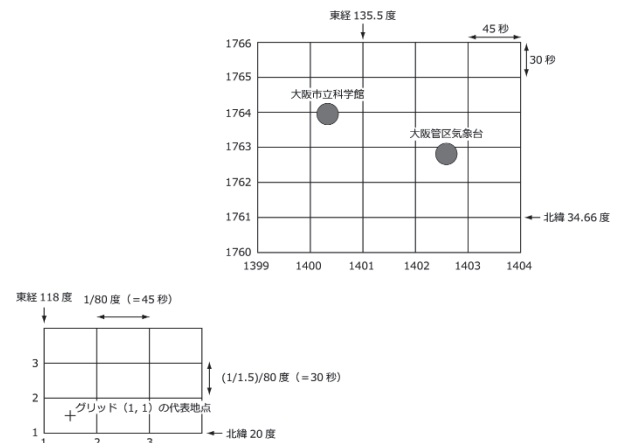


図 2 解析雨量データのグリッドポイント

科学館、および大阪管区気象台のグリッドのインデックスの値は、以下の通りである。

大阪市立科学館

(経度, 緯度) = (135.492, 34.691)

- ・ 経度: $((135.492 - 118) / (1/80)) + 1 = 1400$
- ・ 緯度: $((34.691 - 20) / ((1/80) \times (2/3))) + 1 = 1763$

大阪管区気象台

(経度, 緯度) = (135.519, 34.682)

- ・ 経度: $((135.519 - 118) / (1/80)) + 1 = 1402$
- ・ 緯度: $((34.682 - 20) / ((1/80) \times (2/3))) + 1 = 1762$

機械学習を行う際は、科学館の周辺の一部の範囲を切り出して、データを取り出すこととする。

この際の抽出範囲としては、経度は大阪湾～生駒山に相当する 135.2 度～135.8 度、緯度は狭山市～箕面市に相当する 34.46667 度～34.86667 度とした。これにより、48×48 のメッシュデータが得られる。

表2 抽出データの範囲

	経度	緯度
範囲	135.2 度～135.8 度 (135.8 度は含まない)	34.46667 度 ～ 34.86667 度 (34.86667 度は 含まない)
インデックス	1377～1424	1737～1784
メッシュ数	48	48
範囲目安	大阪湾～生駒山	狭山市～箕面市

なお、解析雨量 GRIB2 ファイルは、降水量がランレングス圧縮で西から東向きに、北から順に格納されている。そのため、python のスクリプトを使ってバイナリデータを直接読み出す際は、配列のインデックスに注意が必要である。

つまり、`rain[0:3360, 0:2560]` という配列に雨量データを取り込んだ場合、先に求めた経度・緯度のインデックスを

緯度: 3360 - 緯度のインデックス
経度: 経度のインデックス - 1

と変換する必要がある。Python の配列は 0 から始まり、最後のインデックスは含まないことにも注意が必要である。

3-2. アメダス降水量との比較

抽出した解析雨量データの妥当性を検証するため、解析雨量データと、アメダス降水量データの比較を行った。

気象庁ホームページより、python のスクリプトを組んで、大阪のアメダスの 1 時間ごとの降水量を取得し、csv ファイルに出力した[4]。

また、解析雨量データから、大阪管区気象台の存在するグリッドポイントの雨量データを抽出した。

2018 年 1 年間の、両者の値を比較したグラフは図 3 の通りである。線形性がとれていることから、正しく抽出されていると言える。

なお、科学館と大阪管区気象台のグリッドポイントの解析雨量データを比較したグラフは図 4 のようになる。2 地点は距離にして、約 2,800m 離れているが、両者の値は、比較的近い値になっていることが分かる。以下の解析において、科学館での雲画像と、大阪アメダスのデータを合わせた解析を行っているが、両者の位置の違いの影響は小さいと考えられる。

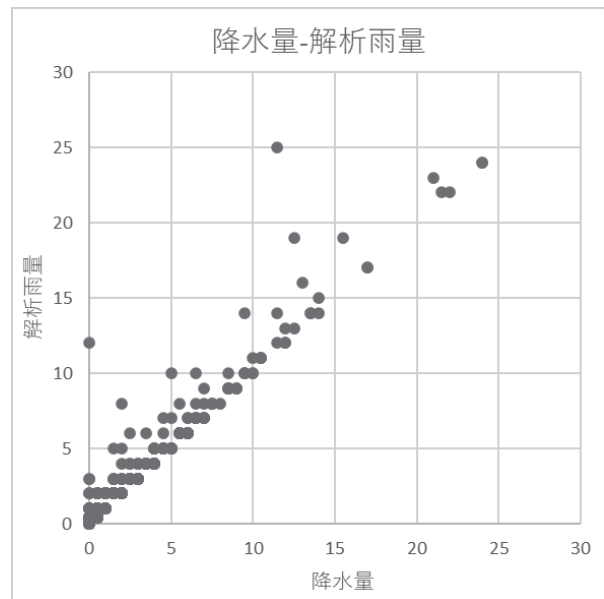


図3 アメダス降水量と解析雨量の関係

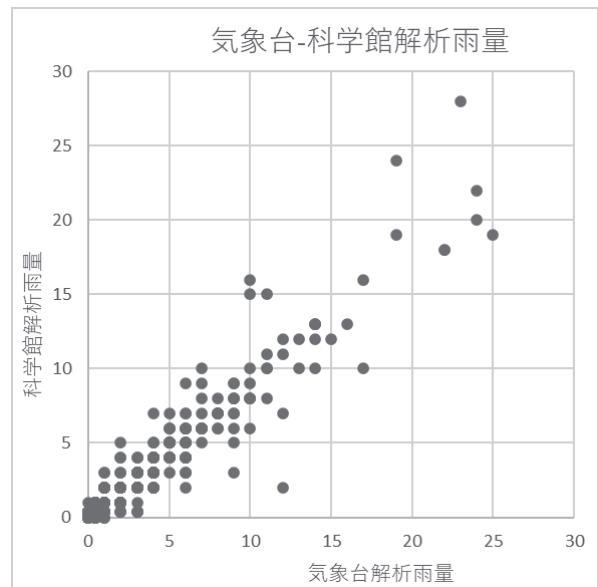


図4 科学館と気象台での解析雨量の関係

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全天画像データである。元データは画像サイズ 640×480 ピクセルの RGB カラーデータであるが、これを 48×48 ピクセルに縮小している。

これに合わせて入力する解析雨量データは、3-1 項で示したように、48×48 の数値データとなっている。

さらに、別の気象データも入力データに加えることを試みた。ここで用いたのは、3-2 項と同じ、気象庁ホームページに公開されている、「大阪(大阪市中央区)」の観測地の過去の気象データで、表 3 に示す項目の通りである。ただし、過去の気象データのうち、「天気」と「雲量」は次項で述べるように教師データとして用いているので、この 2 つの項目は訓練データから省いた。

表3 訓練に用いたデータの項目(気象)

日付	時刻	気圧	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

さらに、各画像データの時刻における、太陽と月の方位角・高度と、月の方位角・高度・位相の天文データを別途計算して、気象データに加えた。これは、画像中に映っている太陽、もしくは月の光が、学習の妨げになっている可能性を考慮したためである。

表4 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 2018 年の 1 年間で、画像枚数はデータの取得できた 1,809 枚を用いた。

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阪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

前報の機械学習モデルを作成する際は、入力データは雲画像のみであったため、プログラミングとしては、単純にモデル層を積み重ねた Sequential API と呼ばれる手法を用いることができた。

表5 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	解析雨量
入力層	画像データを 1 列に	6,912	2,304
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 1,024	1,024	
出力層	ニューロン数 11 or 8	11 または 8	

今回は、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、機械学習モデルを作成する際には複数の入力可能な Functional API を用いてモデルを作成した。

機械学習の手法としては、表 5 に示すニューラルネットワークモデルを用いて、全天雲画像を含めた入力データから、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かめるようにした。

5. 解析結果

5-1. 解析雨量データから気象台の雨量の推定

最初に解析雨量のメッシュデータを入力データとし、気象台の位置での雨量を教師データとして、ニューラルネットワークを用いて機械学習を行ってみた。損失関数の変化を図 5 に、分類結果を図 6 に示す。

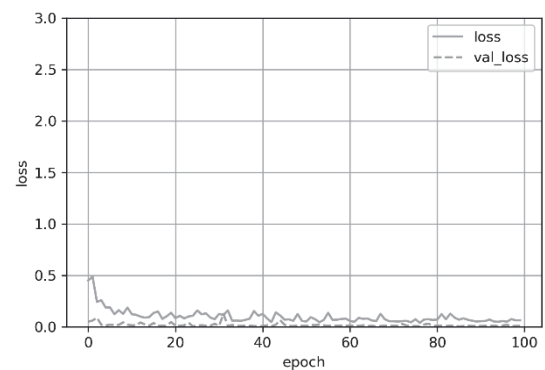


図5 損失関数の変化(ニューラルネットワークモデル)

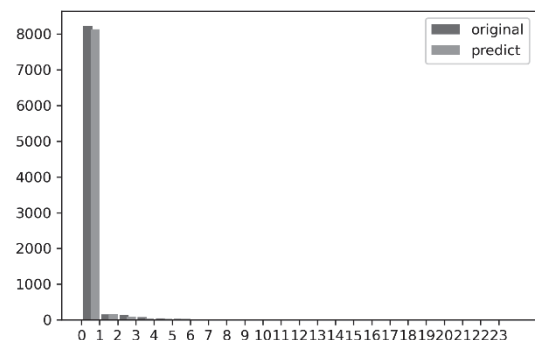


図6 雨量の分類(ニューラルネットワークモデル)

入力データ数は、2018 年の 1 時間ごとのデータであり、計 8,760 である。

本来、メッシュデータの特定の点を選べば正解となる訳であるが、100%ではないものの、高い正答率が得られている。ただし、必ずしも学習に成功するわけでは

く、ここでは上手く学習できたものを例として示している。

なお、図 6 にある通り、8,760 ある入力データのうち、8,000 以上が雨量 1 以下であり、解析雨量データの大部分は、ほとんど降水がない状態であったことが分かる。

5-2. 雲画像データのみによる学習

以下、雲画像と解析雨量データを組み合わせた結果について報告するが、比較のため、前報でも報告した、雲画像のみを入力データとして、ニューラルネットワークモデルを使って雲量の分類を試みた結果を、図 7、図 8 に示す。

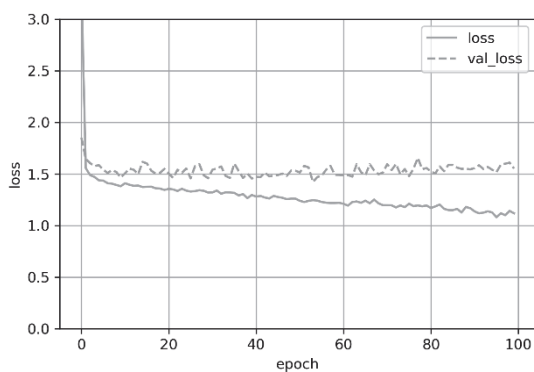


図 7 損失関数の変化(雲画像のみで学習)

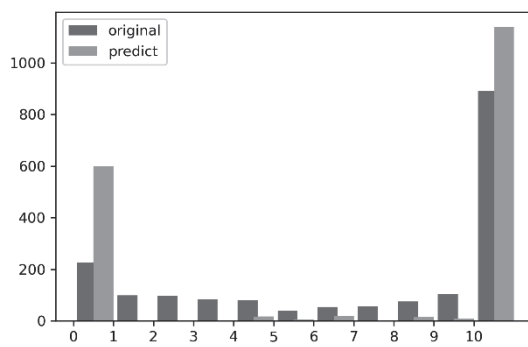


図 8 雲量の分類(雲画像のみで学習)

図 7 は損失関数の変化を示したものである。比較的早い段階で、損失関数の減少が止まっており、長時間学習を進めても、あまり変化は見られない。図 8 の雲量の分類した結果は、ほとんど雲量 0、もしくは 10 に偏ってしまっている。最終的な正答率は、学習用データ 60%、テストデータは 51% であった。

5-3. 雲画像と解析雨量データによる学習

次に、雲画像に追加して、科学館周辺の降雨状況を示す解析雨量のデータを合わせて、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を分類して雲量の分類を試みた結果を、図 9、図 10 に示す。

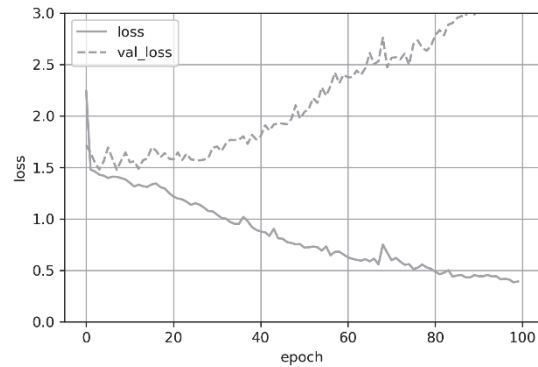


図 9 損失関数の変化(解析雨量データを追加)

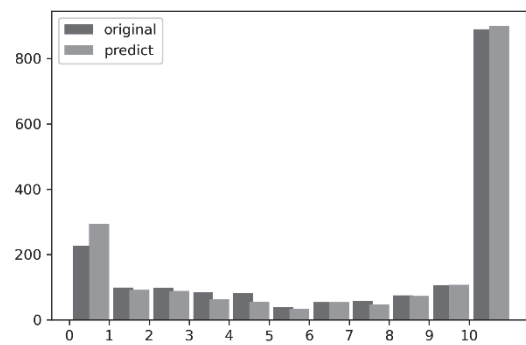


図 10 雲量の分類(解析雨量データを追加)

解析雨量のデータと組み合わせることで、ニューラルネットワークモデルであっても損失関数が減少し、雲量の分類についても、0、10 だけでなく、中間の雲量も適切に分類できるようになった。ただし、テストデータの損失関数は増加しており、過学習となっている。

最終的な正答率は、学習用データ 88%、テストデータは 45% であった。

5-4. 雲画像とアメダス観測値による学習

次に、雲画像に追加して、気象台のアメダス観測による大阪の観測地の気象データを合わせて、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を

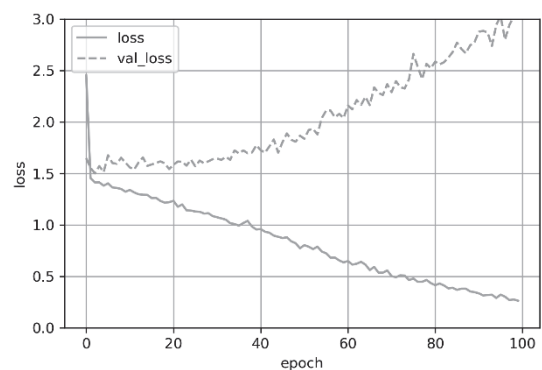


図 9 損失関数の変化(アメダス観測値データを追加)

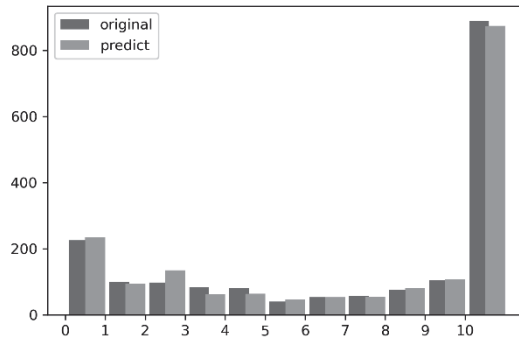


図 10 雲量の分類(アメダス観測値データを追加)

分類して雲量の分類を試みた結果を、図 11、図 12 に示す。

こちら、気象台の観測データと組み合わせることで、ニューラルネットワークモデルであっても損失関数が減少し、雲量の分類についても、適切に分類できるようになった。しかし、やはりテストデータの損失関数は増加しており、過学習となっている。

最終的な正答率は、学習用データ 90%、テストデータは 50%であった。

以上より、雲画像に加えて、解析雨量データや気象台のアメダス観測値を追加して、特徴量を組み合わせると、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となった。しかし、学習用データの損失関数を減少させると、テストデータの損失関数が上昇してしまうという結果になった。雲画像の適切な分類には、さらに別の工夫が必要と考えられる。

6. SHAP 値による分析

雲画像だけでなく、気象データを追加して機械学習を行うことで正解率の向上が見られたことから、特徴量を追加することが、より正確な分類のために有用と考えられる。

そこで、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 11 は雲画像とアメダス観測値を組み合わせると学習させた際に、雲量を 0 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。日照時間や太陽高度、全天日射量が大きく寄与していることから、太陽の影響が大きいことが分かる。一方、ここでは図を示さないが、夜間の雲画像の分類した場合、これらの要素の寄与が小さくなることから、時間帯に応じた適切な特徴量の追加が有用と考えられる。

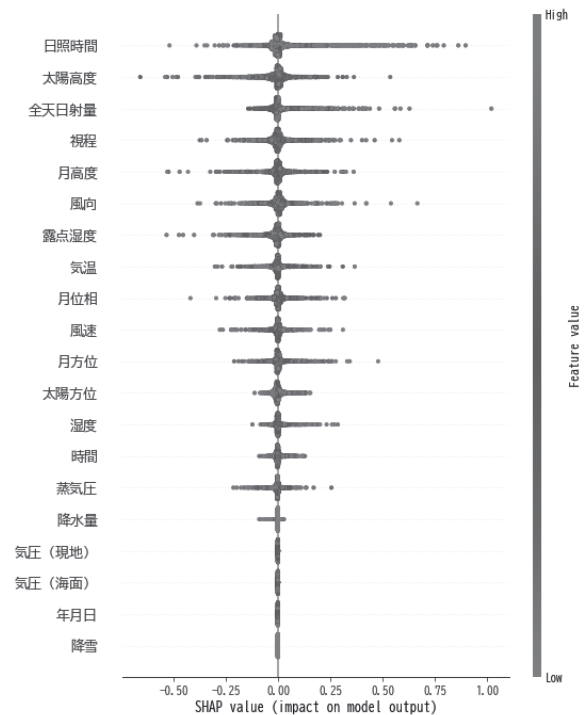


図 11 SHAP 値による分析

7. まとめ

今回、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。

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、入力データとして雲画像のみでは適切な分類ができなかったが、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、より適切な分類が可能となった。一方で、学習用データに対する過学習が起こっているため、未知の画像に対しては、まだ十分な精度を得ることができていない。

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。そのため、時間帯に応じて適切な特徴量を追加することが、より適切な分類に有用と考えられる。

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]気象庁 数値予報解説資料, 51, p106-116(2018)
- [2]江越 航, 大阪市立科学館研究報告, 33, p1-4 (2023)
- [3]気象庁解析雨量(GRIB2 形式)を Python で扱う <https://qiita.com/vpcf/items/b680f504cfe8b6a64222>
- [4]気象庁 過去の気象データ検索 <http://www.data.jma.go.jp/obd/stats/etrn/indwx.php>