

機械学習による星図データに基づく星座識別手法の提案

1 はじめに

近年、機械学習の急速な発展を背景にして、人工知能 (Artificial Intelligence : AI) を用いた認識技術が注目を集めている。特に画像認識の分野においては研究が盛んに行われ、ニューラルネットワークの登場により多くの場合において高い精度での物体の認識が可能となった。一方で、天体写真は背景と対象の線引きが難しく、また撮り手や機材の違いによって撮り方や編集方法が異なるため、そこに写っているものを認識するのは AI をもってしても非常に困難な課題となっている。以上の点を踏まえて本実験では、天体写真を人間のよう認識する AI の構築を最終目的とする研究の初期段階として、天体写真に写っている星座を複数の候補の中から見つけ出す識別器の作成を目的とする。実験の具体的な手順としては、まず訓練データの拡充のために、各星座においてできるだけ写真に似せた星図を作成する。その後作成した星図を畳み込み自己符号化器 (Convolutional Auto Encoder : CAE) を用いて学習させ、学習させた CAE を用いて星図と天体写真の分散表現をそれぞれ獲得する。これにより得られた分散表現をいくつかの手法に適用させることで天体写真に写る星座の識別について数値実験をした。

2 要素技術

2.1 畳み込み自己符号化器

情報量を落とさずにより低次元の特徴表現を獲得するためのニューラルネットワーク手法に自己符号化器 (Auto Encoder : AE) [1] がある。AE は出力されるデータが入力されたデータと一致するように損失関数を定め、学習を進める。また AE は、中間層出力を取り出すことで、データの性質を反映した分散表現が得られるという特徴がある。この AE に対して畳み込み層を追加したものを畳み込み自己符号化器 (Convolutional Auto Encoder : CAE) という。CAE では、エンコーダ部に畳み込み層と MaxPooling 層を、デコーダ部に畳み込み層と UpSampling 層を用いている。なお本実験では、天体写真を合理的に実数値ベクトル化するために用いた。

- | | | |
|---------|--------|--------|
| ・はくちょう座 | ・ふたご座 | ・いて座 |
| ・カシオペア座 | ・こぐま座 | ・こと座 |
| ・おおぐま座 | ・おおいて座 | ・オリオン座 |
| ・おとめ座 | ・おうし座 | ・ペガサス座 |
| ・さそり座 | ・しし座 | ・わし座 |

図 1: 本実験で使用した星座

2.2 サポートベクターマシン

サポートベクターマシン (Support Vector Machine : SVM) [2] は、パターン識別用の教師あり機械学習方法である。マージン最大化という考えに基づき分類するため、二クラス分類問題に非常に優れているだけでなく、カーネルトリックという手法を使うことで線形分離不可能な場合においても適用が可能であるので応用範囲も広い。また二クラス分類を行う分類器を複数組み合わせることで多クラス分類を実現する One-versus-the-rest [3] や One-versus-one という手法も存在する。なお本実験では One-versus-the-rest を使用して天体写真に写る星座の識別をした。

2.3 コサイン類似度

ベクトル空間モデルにおいて二本のベクトルがどれほど類似しているかの距離指標としてコサイン類似度がある。ベクトル同士のなす角度の近さを表現するため、1 に近ければ類似しており、0 に近ければ類似していないということになる。なお本実験では星図と天体写真の分散表現ベクトルの類似度を調べるのに使用した。

3 実験

3.1 データセット

本実験では比較的ポピュラーな 15 種類の星座を対象とした。図 1 に対象とした星座名を示す。訓練画像には 3.2 節で示す方法を用いて作成した合計 45 枚の星図を使用した。またテスト画像には各星座につきそれぞれ 1 枚ずつ、合計 15 枚を Web からスクレイピングしたものを使用した。なおこれらの画像

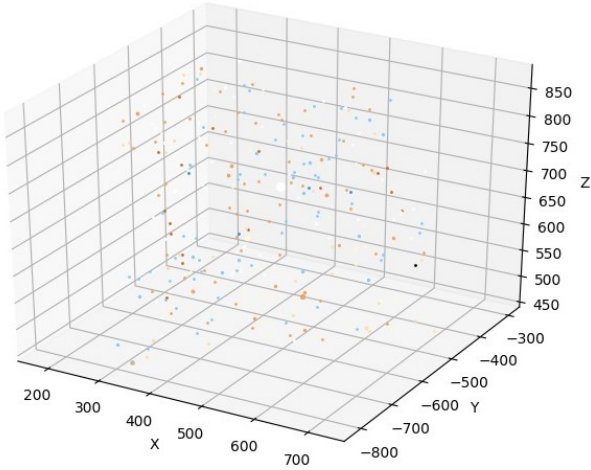


図 2: はくちょう座の三次元空間のプロット図



図 3: 完成したはくちょう座の星図

には, 15 種類の星座名をラベルとして与えた. また今回は実験の簡易化のため, テスト画像の天体写真は, できる限り星が大きく写っているものかつ大きさ以外の向きについては星図と同じ構図であるものを使用した.

3.2 星図の作成

まず星座を構成する星の位置, 大きさ, 色の情報を加味した星図を作成する.

図 1 に示した星座に関して, 構成する星の赤経, 赤緯, 視等級, スペクトル型を Wikipedia よりそれぞれクロールして獲得した. ここで, 赤経と赤緯は地平座標に変換する必要がある. 通常赤経は $A[h] B[m] C[s]$, 赤緯は $D^\circ E' F''$ のような形で与えられ, ラジアン表記の赤経 α と赤緯 δ は以下の式 1, 2 から得られる.

$$\alpha = (A + (B + C/60)/60) \times 15 \times \pi/180, \quad (1)$$



図 4: フィルタ処理をしたはくちょう座の星図

$$\delta = \begin{cases} (D + (E + F/60)/60) \times \pi/180 & (D \geq 0) \\ -(D + (E + F/60)/60) \times \pi/180 & (D < 0). \end{cases} \quad (2)$$

得られた赤経 α と赤緯 δ から以下の 3 ~ 5 式を用いて地平座標に変換される. なお r は投影天球の半径であり, 本実験では 1000 とした. また x 軸は春分点の方向, z 軸は天の北極の方向である.

$$x = r \cos \alpha \cos \delta, \quad (3)$$

$$y = r \sin \alpha \cos \delta, \quad (4)$$

$$z = r \sin \delta. \quad (5)$$

これによって計算された地平座標を基に, 三次元空間上に各星をプロットした. このとき, 投影天球の半径 1000 に対し星の直径は $20/(\text{視等級})$ とし, 星の色はスペクトル型を反映したものとする. 図 2 に例としてはくちょう座の三次元空間のプロット図を示す. この空間に対して, 任意の平面を定め, そこにすべての星を投影させる. なおこの平面は各星座において原点との位置関係によりそれぞれ独自で決定したものであり, 平面自体は一面が黒一色であるものである. 最後に, 星を投影した平面に対してガウスフィルタをかけて星図を完成させた. 図 3 に完成したはくちょう座の星図を示す. また, 完成した星図に対して,

フィルタ 1: 画像内で最も大きな星の面積の $1/30$ よりも小さな面積をもつ星を削除するフィルタ

フィルタ 2: 全体的に縮小処理をかける

の二種類のフィルタをかけたものを追加で生成した. 図 4 にフィルタ処理をした結果を示す. このフィルタ処理により, 星の数が多い写真から少ない写真まで幅広く対応できるようにした. 同じことを残りの 14 星座に対しても適用し, 合計で 45 枚の星図を完成させた.

表 1: CAE の学習パラメータ

| 学習アルゴリズム | Adam |
|----------|----------------------|
| 学習率 | 1.0×10^{-5} |
| 損失関数 | binary crossentropy |
| エポック数 | 100 |
| バッチサイズ | 16 |

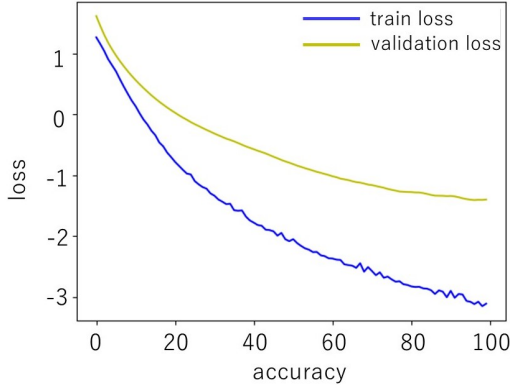


図 5: loss の推移

3.3 CAE の学習および分散表現の獲得

今回は深層学習ライブラリの keras を用いて, 作成した 45 枚の星図を CAE で学習させた. このとき, 星図のサイズは 128×128 に統一した. 表 1 に CAE の学習パラメータを示す. なお中間層のサイズは 8,192 であった.

図 5 に loss の推移を示す. これにより, 学習は進んでいると考えられるため星図データを学習した CAE を作成することができた. したがって今度はこの CAE を用いて, 星図 45 枚とテスト画像の天体写真 15 枚の分散表現を獲得した. なおテスト画像の天体写真に関しては撮り手の編集の都合により背景が白飛びしているものもあることを考慮して, 画像の輝度に関して閾値を設け, それよりも小さな輝度をもつピクセルは黒塗りをする処理を施した.

3.4 天体写真の識別

3.4.1 SVM を用いた識別

今回は機械学習ライブラリの scikit-learn を用いて, 得られた分散表現を SVM に適用することで天体写真の識別を試みた. 今回は多クラス分類の代表として, One-versus-the-rest を採用した. なおパラメー

表 2: SVM の最良パラメータ

| kernel | rbf |
|----------|----------------------|
| C | 1.5127 |
| γ | 1.6×10^{-5} |

タは Optuna を用いて最適化した. 表 2 に Optuna によって得られた最良パラメータを示す. また表 3 に SVM で識別をしたときの結果を示す. SVM が写真を識別した星座に 1 が, それ以外には 0 が記載されている. なお正しく 1 が導かれたものを橙色に, 間違っ 1 が導かれたものを青色で表示している. 全体的に高精度に識別できたが, はくちょう座やおおぐま座といった, 人間の目で見ても比較的わかりやすい星座をうまく識別できなかったことに関して今後原因を明らかにする必要がある.

3.4.2 コサイン類似度を用いた識別

次に 2.3 節で示したコサイン類似度を使用して天体写真の識別を試みた. 各写真と各星図との類似度は, 写真と 3 種類ある各星図の類似度をそれぞれ計算し, 最も類似度が高いものが選択されている. 表 4 に結果を示す. 各星座において最も類似度が高かったものを着色しており, 正しく識別できたものを橙色で, 間違っ 1 が導かれたものを青色で表示している. コサイン類似度を用いた識別でも SVM を用いたときと同様に高い精度で識別できた. また SVM では識別できなかったはくちょう座やいて座, ペガサス座も識別できている. これは, 訓練画像の枚数が少なかったために不要な学習を数多くするよりも直接類似度を調べた方が効率的であったと考えられる.

4 まとめと今後の課題

今回の実験により, 作成した星図を使用することで高い精度で天体写真に写る星座が識別ができた. 特にコサイン類似度を使用する識別法では正答率が 80% という高い精度の結果が得られた. 以上の結果より星図を使用した星座識別の有効性を確認できたといえる. 一方でこぐま座やおおぐま座はどちらの識別手法でもうまく識別できなかったため, 今後は

表 3: SVM の識別結果

| | | 星図 | | | | | | | | | | | | | | |
|--------|-------|-------|-----|----|-------|-----|----|------|------|------|-----|-----|------|-----|----|----|
| | | はくちょう | ふたご | いて | カシオペア | こぐま | こと | おおぐま | おおいぬ | オリオン | おとめ | おうし | ペガサス | さそり | しし | わし |
| 写真 | はくちょう | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | ふたご | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | いて | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | カシオペア | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | こぐま | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | こと | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | おおぐま | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | おおいぬ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | オリオン | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | おとめ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | おうし | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| | ペガサス | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| | さそり | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| | しし | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | わし | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 正答率 | | 0.667 | | | | | | | | | | | | | | |
| ベースライン | | 0.067 | | | | | | | | | | | | | | |

表 4: コサイン類似度の計算結果

| | | 星図 | | | | | | | | | | | | | | |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | はくちょう | ふたご | いて | カシオペア | こぐま | こと | おおぐま | おおいぬ | オリオン | おとめ | おうし | ペガサス | さそり | しし | わし |
| 写真 | はくちょう | 0.419 | 0.260 | 0.327 | 0.332 | 0.201 | 0.180 | 0.294 | 0.187 | 0.218 | 0.306 | 0.407 | 0.255 | 0.350 | 0.197 | 0.195 |
| | ふたご | 0.329 | 0.427 | 0.397 | 0.371 | 0.235 | 0.298 | 0.307 | 0.343 | 0.374 | 0.261 | 0.359 | 0.357 | 0.310 | 0.328 | 0.226 |
| | いて | 0.471 | 0.258 | 0.474 | 0.276 | 0.241 | 0.303 | 0.283 | 0.247 | 0.243 | 0.241 | 0.381 | 0.274 | 0.200 | 0.215 | 0.262 |
| | カシオペア | 0.484 | 0.428 | 0.477 | 0.531 | 0.363 | 0.368 | 0.419 | 0.411 | 0.450 | 0.473 | 0.457 | 0.447 | 0.440 | 0.419 | 0.296 |
| | こぐま | 0.382 | 0.326 | 0.299 | 0.441 | 0.362 | 0.311 | 0.271 | 0.333 | 0.377 | 0.366 | 0.250 | 0.305 | 0.190 | 0.246 | 0.308 |
| | こと | 0.262 | 0.148 | 0.242 | 0.157 | 0.154 | 0.773 | 0.310 | 0.484 | 0.273 | 0.119 | 0.343 | 0.141 | 0.170 | 0.345 | 0.190 |
| | おおぐま | 0.371 | 0.398 | 0.276 | 0.258 | 0.269 | 0.318 | 0.278 | 0.304 | 0.281 | 0.294 | 0.319 | 0.367 | 0.241 | 0.363 | 0.312 |
| | おおいぬ | 0.314 | 0.200 | 0.245 | 0.264 | 0.272 | 0.215 | 0.309 | 0.445 | 0.297 | 0.376 | 0.255 | 0.329 | 0.200 | 0.310 | 0.215 |
| | オリオン | 0.314 | 0.249 | 0.319 | 0.371 | 0.208 | 0.302 | 0.241 | 0.352 | 0.433 | 0.337 | 0.278 | 0.273 | 0.189 | 0.176 | 0.255 |
| | おとめ | 0.245 | 0.202 | 0.313 | 0.274 | 0.232 | 0.162 | 0.387 | 0.226 | 0.226 | 0.396 | 0.181 | 0.304 | 0.280 | 0.283 | 0.190 |
| | おうし | 0.452 | 0.381 | 0.392 | 0.410 | 0.319 | 0.324 | 0.403 | 0.357 | 0.353 | 0.413 | 0.450 | 0.418 | 0.366 | 0.364 | 0.267 |
| | ペガサス | 0.226 | 0.194 | 0.231 | 0.244 | 0.188 | 0.170 | 0.267 | 0.178 | 0.184 | 0.317 | 0.286 | 0.321 | 0.213 | 0.256 | 0.198 |
| | さそり | 0.278 | 0.269 | 0.291 | 0.265 | 0.245 | 0.181 | 0.335 | 0.271 | 0.211 | 0.333 | 0.189 | 0.301 | 0.501 | 0.274 | 0.177 |
| | しし | 0.275 | 0.175 | 0.234 | 0.449 | 0.217 | 0.169 | 0.375 | 0.176 | 0.187 | 0.279 | 0.224 | 0.297 | 0.318 | 0.626 | 0.193 |
| | わし | 0.299 | 0.178 | 0.280 | 0.250 | 0.245 | 0.349 | 0.359 | 0.342 | 0.247 | 0.281 | 0.271 | 0.290 | 0.189 | 0.318 | 0.391 |
| 正答率 | | 0.800 | | | | | | | | | | | | | | |
| ベースライン | | 0.067 | | | | | | | | | | | | | | |

その原因究明と解決が課題である。また今回は星図と天体写真の構図をそろえて実験をしたが、今後は回転画像を追加するなどして構図が異なる星座に関しても識別が可能かどうか検証していく予定である。加えて今回は比較的星が大きく写った天体写真を識別対象としたため、今後はそのような条件を設けないより一般的な天体写真も識別できる方法を検討していく。

参考文献

- [1] 中田 秀基 and 麻生 英樹. Variational autoencoder を用いた画像スタイル変換 (バイオメトリクス).

電子情報通信学会技術研究報告 = *IEICE technical report* : 信学技報, 117(513):121–126, mar 2018.

- [2] 津田 宏治. サポートベクターマシンとは何か. 電子情報通信学会誌, 83(6):460–466, jun 2000.
- [3] 小林 卓夫 and 清水 郁子. One-versus-the-rest 法における多クラス分類の識別境界の考察. 電子情報通信学会技術研究報告. *PRMU*, パターン認識・メディア理解, 109(344):65–70, dec 2009.