

知能システム論 第 3 回レポート

阿部 桃大

2023 年 10 月 25 日

1 概要

今回の課題ではポケモンの種族値のデータベースを用いて、PCA の後に k-means を行い、3 つのクラスタに分類する解析を行った。また、ガウシアンカーネルを用いて同様に PCA の後に k-means を行い、3 つのクラスタに分類する解析を行った。

なお、解析元となるデータは、<https://www.kaggle.com/datasets/abcsds/pokemon> から取得した。

2 実装

本課題の実装は、`kadai.py` にて行った。このプログラムは、`Pokemon.csv` からポケモンの 6 つの種族値 (H,A,B,C,D,S) を読み込み、上記の解析を行う。

コードの詳細については添付のコード内にコメントとして記載している。

3 解析結果

以下に 2 つの手法を用いて行った解析の結果を示す。

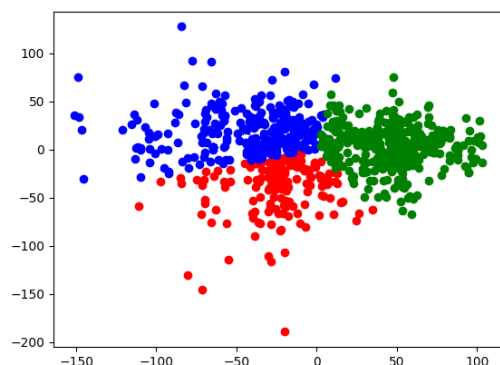


図 1 PCA の後に k-means を行った結果

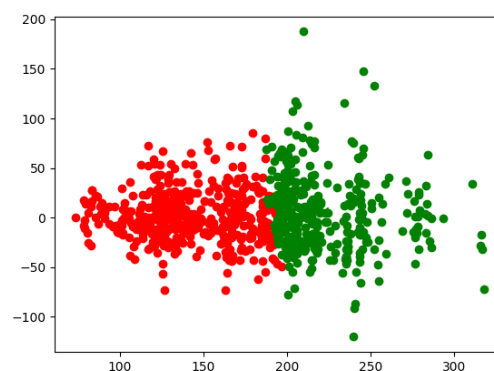


図 2 ガウシアンカーネルを用いて PCA の後に k-means を行った結果

4 考察

通常の PCA 及び k-means を用いた場合、図 1 のように、3 つのクラスタに分類された。しかし、各クラスタの境界には多くのデータが含まれており、クラスタの分類がうまくいっていないと考えられる。

一方で、ガウシアンカーネルを用いた PCA 及び k-means を用いた場合、図 2 のように、2 つのクラスタに分類された。この結果においても、各クラスタの境界には多くのデータが含まれており、クラスタの分類がうまくいっていないと考えられる。ここで、本来は 3 つのクラスタに分類されるはずのデータが 2 つのクラスタに分類されたことから、ガウシアンフィルタを用いるとこのデータが本質的に 2 つのクラスタに分類されるような特性を持つものだと考えられる。

最後に各クラスタの意味についても軽く考察しておく。

通常の PCA 及び k-means を用いた場合、クラスタ 1 は HP が高く、クラスタ 2 は HP が低いという特徴が見られる。また、ガウシアンカーネルを用いた PCA 及び k-means を用いた場合、x 軸は $(-0.30, -0.49, -0.38, -0.51, -0.39, -0.32)$ の固有ベクトルに対応し、y 軸は $(-0.04, -0.07, -0.70, 0.38, -0.17, 0.58)$ の固有ベクトルに対応する。したがって x が高いポケモンは全体的に種族値が低い傾向にあり、未進化ポケモンなどが多く該当すると思われる。y が高いポケモンは特攻が低く、防御と素早さが高い傾向にある。つまり、緑色は未進化ポケモンなどの低種族値ポケモンで、青色は特攻が低く、防御と素早さが高いポケモン、逆に赤色は特攻が高く、防御と素早さが低いポケモンと考

えられる。

ガウシアンカーネルを用いた手法のほうでは二つのクラスタを分けているのは x 成分であるが、これも上の場合と同様に $(0.43, 0.38, 0.38, 0.41, 0.45, 0.38)$ の固有ベクトルに対応し、全体的な種族値を表すパラメータであるといえる。すなわち、赤色は全体的に種族値が低い未進化などのポケモン、緑色は全体的に種族値が高い伝説ポケモンや最終進化ポケモンなどと考えられる。