寶可夢數值預測

11377034 資訊碩一 姜柏仰

1. 資料集

名稱: Pokemon with stats and image

欄位: Image, Index, Name, Type1, Type2, Total, HP, Attack, Defense, SP. ATK. , SP.DEF , Speed

1. 前處理
   * 將Type1與Type2進行one hot encoding
   * 新增Legendary欄位 ( 將傳說寶可夢標註為1，非傳說寶可夢標註為0 )
2. 目標
   * 利用寶可夢的特徵預測其種族值( HP, Attack, Defense, SP. ATK. , SP.DEF , Speed )
   * 使用的欄位: 英文名字、圖片、是否為傳說寶可夢、屬性
3. 方法
   * 模型: XGBoost、Resnet50
   * 步驟
     1. 將英文名字用Label Encoding進行編碼
     2. 將圖片使用Resnet50提取特徵
     3. 利用Grid Search尋找最佳超參數
     4. 利用XGBoost預測種族值
4. 結果
   * 評估指標: RMSE( Root Mean Square Error )

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Index | HP | Attack | Defense | SP. ATK. | SP. DEF. | Speed |
| 1 | 22.63 | 30.64 | 28.78 | 28.80 | 25.82 | 28.42 |
| 2 | 22.57 | 30.48 | 28.21 | 27.77 | 25.57 | 28.43 |
| 3 | 17.95 | 26.15 | 24.97 | 25.17 | 23.45 | 26.16 |

* + 其中Index為我試過的3種方法，分別為1. 單純使用XGBoost。 2. 加入Grid Search進行超參數最佳化。 3. 再加入Resnet50對圖片進行特徵擷取，上表證明在這三種方法中RMSE最小效果最好
  + 除此之外還額外分出18種屬性並且分別計算%RMSE得出下表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| %RMSE(%) | HP | Attack | Defense | SP. ATK. | SP. DEF. | Speed |
| Normal | 22.18 | 22.50 | 13.89 | 15.37 | 17.88 | 23.86 |
| Fighting | 17.07 | 23.80 | 22.02 | 19.23 | 21.04 | 20.28 |
| Flying | 14.86 | 18.27 | 18.53 | 21.16 | 22.26 | 15.05 |
| Poison | 16.50 | 20.30 | 20.75 | 20.46 | 18.71 | 20.20 |
| Ground | 21.16 | 19.53 | 19.75 | 14.29 | 16.03 | 16.67 |
| Rock | 14.79 | 23.41 | 33.01 | 23.28 | 24.10 | 21.96 |
| Bug | 12.27 | 27.27 | 24.54 | 19.82 | 20.56 | 16.09 |
| Ghost | 11.80 | 22.36 | 25.75 | 24.04 | 21.09 | 30.31 |
| Steel | 17.24 | 21.74 | 19.51 | 21.99 | 27.64 | 18.19 |
| Water | 18.44 | 24.78 | 22.19 | 18.66 | 22.94 | 20.61 |
| Fire | 13.99 | 16.44 | 15.30 | 23.04 | 18.24 | 19.72 |
| Grass | 13.79 | 17.47 | 16.47 | 18.47 | 11.96 | 20.84 |
| Electric | 17.07 | 22.21 | 17.22 | 25.17 | 16.55 | 28.37 |
| Psychic | 21.78 | 23.14 | 20.06 | 29.85 | 17.49 | 24.85 |
| Ice | 20.60 | 16.45 | 30.96 | 27.70 | 21.75 | 32.71 |
| Dragon | 22.38 | 14.35 | 17.70 | 26.66 | 14.51 | 24.22 |
| Dark | 22.46 | 14.23 | 16.38 | 24.58 | 20.15 | 17.05 |
| Fairy | 16.97 | 17.64 | 17.48 | 23.29 | 25.01 | 24.22 |

* + 綜合而言HP的預測最好%RMSE來到平均17.57%
  + Speed的預測最差%RMSE來到平均21.65%