



AI CUP 期末報告

613430062 陳姿羽

613430008 林建君

613430015 石明諺

613430058 蔡欣宏

613430067 詹秉融



特徵與目標變數處理

1. 特徵：

- 每筆揮拍資料從 txt 檔切分為 27 段後，提取以下特徵
 - 加速度與角速度的平均值、變異數、RMS、最大/最小值
 - 傅立葉頻譜特徵
 - 峰度、偏度、譜熵
 - 共 34 維特徵，搭配 FFT 輔助提取時域與頻域資訊

2. 目標變數：

- gender（二元分類）
- hold racket handed（二元分類）
- play years（多類別分類，共 3 類）
- level（多類別分類，共 4 類，從 2 到 5）

資料切分方式 & 特徵縮放

1. 資料切分方式

- 根據 `player_id` 進行切分（`train_test_split()`），確保相同的受測者不會同時出現在訓練與驗證中

2. 特徵縮放

- 使用 `StandardScaler` 將特徵標準化為 0 均值、1 標準差。
- 為後續 CNN、LSTM 與 Transformer 訓練創造穩定分布。

模型方法

1. 模型架構

- 二元分類任務使用 CatBoostClassifier (depth=2)
- 多分類任務使用 CNN → BiLSTM → Transformer → FC 預測層

2. 特點:

- gender 與 hold 使用 CatBoost 作為 baseline 高效分類器。
- play years 與 level 使用自建深度神經網路 (CNN + BiLSTM + Transformer 架構)
- 每次測驗視為一段序列 (27 揮拍 × 34 特徵)，模型能捕捉時間動態資訊

3. 模型參數

- BiLSTM 模型以 batch size 64、學習率 0.0005 訓練 120 次，採用 Adam 最佳化器與 CrossEntropyLoss。

預測流程概述

1. `main()`：訓練模型並對驗證資料進行預測
 - 資料整理 → 特徵縮放 → 分群建模
 - 訓練四個分類器（2 個CatBoost + 2 個BiLSTM）
 - 輸出機率預測至 `valid_prediction.csv`
2. `predict_new_data()`：對測試資料使用已訓練模型預測
 - 載入 test 資料
 - 使用訓練階段保存的 scaler、CatBoost 模型與 LSTM 權重
 - 輸出 `test_prediction.csv`

實驗結果

```
(base) C:\Users\USER\Desktop\AI_CUP>python aicup_v3.py
執行模型訓練...
Gender AUC: 0.9627
Hold Hand AUC: 1.0000
Play Years AUC: 0.6539
Level AUC: 0.8435
Final Score (Average AUC): 0.8650
預測完成，輸出檔為 valid_prediction.csv
模型訓練完成，是否執行測試？(yes/no): yes
File: tabular_data_test\1968.csv, Data Shape: (27, 34)
File: tabular_data_test\1969.csv, Data Shape: (27, 34)
File: tabular_data_test\1970.csv, Data Shape: (27, 34)
File: tabular_data_test\1971.csv, Data Shape: (27, 34)
File: tabular_data_test\1972.csv, Data Shape: (27, 34)
```

⋮

```
File: tabular_data_test\3407.csv, Data Shape: (27, 34)
File: tabular_data_test\3408.csv, Data Shape: (27, 34)
File: tabular_data_test\3409.csv, Data Shape: (27, 34)
File: tabular_data_test\3410.csv, Data Shape: (27, 34)
File: tabular_data_test\3411.csv, Data Shape: (27, 34)
預測完成，輸出檔為 test_prediction.csv
```

	A	B	C	D	E	F	G	H	I	J
1	unique_id	gender	hold racke	play years	play years	play years	level_2	level_3	level_4	level_5
2	1968	0.98	1	0	1	0	0	1	0	0
3	1969	0.97	1	0	1	0	0	0	0	1
4	1970	0.96	1	0	1	0	0	0	0	1
5	1971	0.97	1	0	1	0	0	0	0	1
6	1972	0.97	1	0	1	0	0	0	0	1
7	1973	0.99	1	0	1	0	0.99	0	0	0.01
8	1974	0.96	1	0	1	0	0	0	0	1
9	1975	0.98	1	0	1	0	0	0	0	1
10	1976	0.98	1	0.12	0.88	0	0	0	0	1
11	1977	0.95	1	0	1	0	0	0	0	1
12	1978	0.97	1	0	1	0	0	0	0	1
13	1979	0.96	0.99	0	1	0	0	0	0	1
14	1980	0.97	0.9	0.02	0.76	0.23	0	0	1	0
15	1981	0.98	0.06	0.01	0.99	0	0	0	0	1
16	1982	0.97	1	0	1	0	0	0	0	1
17	1983	0.97	0.45	0	1	0	0.01	0.34	0	0.65
18	1984	0.98	1	1	0	0	0	0	0	1
19	1985	0.99	0.29	0.04	0.96	0	0	0	0	1
20	1986	0.98	1	0	1	0	0	0	0	1
21	1987	0.97	1	0	1	0	0	0	0	1
22	1988	0.98	1	0	0	1	1	0	0	0
23	1989	0.97	1	0	1	0	0	0	0	1
24	1990	0.98	0.19	0	1	0	0	0	0	1

競賽成績名次



1. Public Leaderboard

222	TEAM_7561	5	20	0.782928	5/30/2025 3:49:39 PM
-----	-----------	---	----	----------	-------------------------

2. Private Leaderboard

245	TEAM_7561	5	20	0.711310	5/31/2025 1:39:37 PM
-----	-----------	---	----	----------	-------------------------

3. Submission

 test_prediction.csv 2025/05/30 v3 加入transformer num_heads=8 上傳成員 蔡 欣宏	2025-05-30 03:49:39	0.78292775	0.7036142872233537	Scoring success.
 test_prediction.csv 2025/05/31 v1 上傳成員 蔡 欣宏	2025-05-31 01:39:37	0.77026644	0.7113109332205685	Scoring success.

GitHub 連結

https://github.com/tzuyu-TYC/AI_CUP.git