

# AI CUP 2025 春季賽

## 桌球智慧球拍資料的精準分析競賽報告

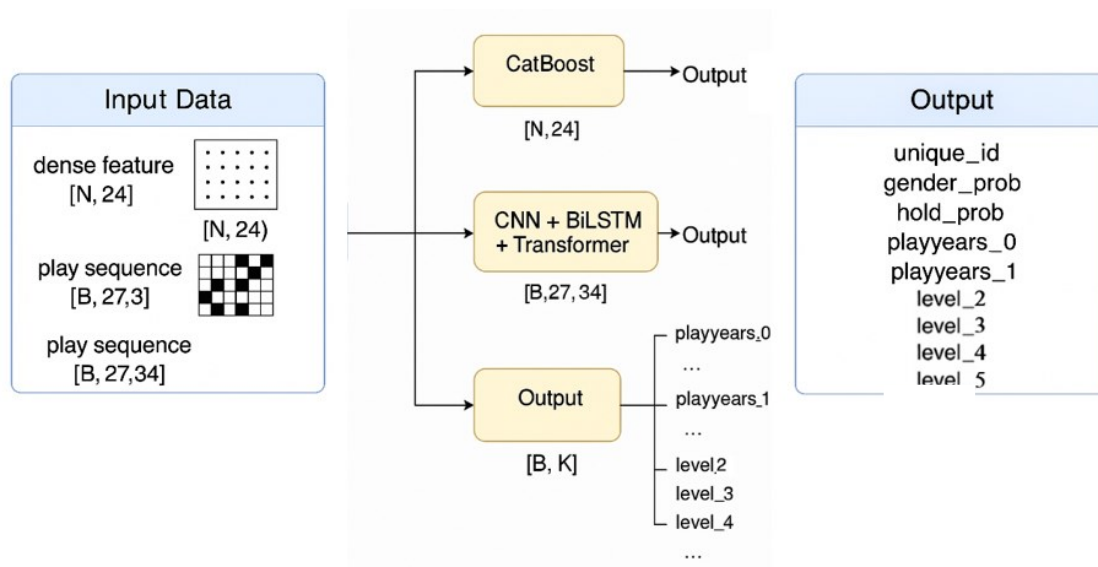
隊伍：TEAM\_7561

隊員：613430062 陳姿羽(隊長)、613430008 林建君、613430015 石明諺、613430058 蔡欣宏、613430067 詹秉融

Private leaderboard：0.711310 / Rank 245

### 壹、環境

這次任務中的開發與訓練皆於 Windows 11 作業系統下進行，主要使用 Python 3.9 作為開發語言。資料處理與建模所使用的函式庫：Pandas、NumPy、SciPy、Matplotlib、Scikit-learn、joblib 等。機器學習模型使用 LightGBM、XGBoost、CatBoost 等分類器，深度學習使用 PyTorch 進行模型架構設計與訓練，包含 CNN、BiLSTM 及 Transformer 模組。資料前處理使用 FFT 來萃取頻域特徵，利用 `scipy.fft`。模型訓練皆在 CPU 環境下執行。訓練模型分別使用兩種，在二元分類的模型使用 CatBoostClassifier。在多分類的模型使用 CNN + BiLSTM + Transformer 混合模型，並訓練於官方所提供的桌球選手揮拍的資料集，以下為輸入與輸出型式：



### 貳、演算法方法與模型架構

本專案的演算法設計依據四個預測目標的任務屬性（Binary Classification / Multi-Class Classification）進行模型選擇與調整。整體系統的處理流程分為「模型輸入

與任務劃分」、「模型建構與訓練」、「模型參數」三個核心模組。

## 1. 模型輸入與任務劃分

輸入特徵為從原始 6 軸感測器資料（加速度  $ax/ay/az$ ，角速度  $gx/gy/gz$ ）提取的統計與頻譜特徵。每一筆測資先均分為 27 段，每段各自計算以下 34 維特徵，包括加速度與角速度的平均值、變異數、RMS、最大/最小值、傅立葉頻譜強度、功率譜密度、峰度、偏度與譜熵，涵蓋時域與頻域資訊。資料標準化採用 StandardScaler，使特徵具有 0 均值與單位變異數，以利深度模型穩定訓練。模型最終的輸入為  $shape = [B, 27, 34]$  的三維張量，表示每位選手一筆測試資料由 27 段揮拍資料組成，每段有 34 維特徵。

## 2. 模型建構與訓練

模型依任務性質設計：對於 gender 與 hold racket handed 等二元分類任務，選用 CatBoostClassifier，使用樹深度為 2、L2 正則項為 5，快速收斂且具備強健性；對於 play years（3 類）與 level（4 類）多分類任務，採用自建 CNN+BiLSTM+Transformer 混合神經網路，採用的原因是具備提取區域特徵、捕捉時序關聯與建模長距依賴三重優勢。其架構先透過 1D CNN 對特徵時間軸進行濾波與壓縮，接續以雙向 LSTM 擷取時間順序關聯，最後經 Transformer 模型強化序列間長距依賴關係，最後由全連接層進行分類預測。

- CNN 模組
  1. 輸入維度  $[B, 27, 34]$  經過轉置後變為  $[B, 34, 27]$
  2. 經由  $1D\ Conv \rightarrow BatchNorm \rightarrow ReLU \rightarrow Dropout (p=0.3)$  進行特徵壓縮與非線性轉換
  3. 輸出  $shape = [B, C, 27]$ ，其中  $C$  為 CNN channels（預設 32）
- BiLSTM
  1. 輸入轉為  $[B, 27, C]$ ，採雙向 LSTM，提取時間序列上的雙向模式
  2.  $hidden\_size = 64$ ，輸出為  $[B, 27, 128]$ （2 倍於  $hidden\_size$ ）
  3. LSTM 層數：2，採  $batch\_first=True$  設計以對齊 PyTorch Tensor 格式
- Transformer Encoder
  1. 將 LSTM 輸出作為 Encoder 輸入，補足長距序列間關聯建模
  2. 採  $nhead=8, dim\_feedforward=128, num\_layers=2$

3. 最終取  $[B, 27, 128]$  的最後一個 timestep，輸入 FC 層
- FC 層
    1. 全連接層輸出  $[B, K]$ ，其中  $K=3$  (play years) 或  $K=4$  (level)
    2. 最後經 softmax 得到各類別機率分布

### 3. 模型參數

BiLSTM 訓練參數包括 batch size 64、學習率 0.0005、訓練輪數為 120，損失函數採 CrossEntropyLoss，優化器則選用 Adam。模型於訓練後輸出至 .pth 檔，並於推論階段呼叫載入。

## 參、創新性

本專案的創新處主要展現在模型設計與預測策略兩個面向。

在模型方面，我們提出一個結合 CNN、BiLSTM 與 Transformer 的多層次時序特徵學習架構，藉以提升模型對於揮拍動作的時序理解能力。首先，CNN 模組負責提取局部時間窗內的高階特徵，能夠有效濃縮短時間內的變化資訊，例如動作啟動瞬間的加速度特徵。接著，BiLSTM 模組則進一步捕捉動作序列的前後文依賴關係，提供雙向的時間上下文資訊，有助於辨識動作中潛藏的時序規律與過渡特徵。而最後引入的 Transformer 結構，透過自注意力機制補足長距離依賴資訊，解決傳統 RNN 類模型在處理長序列時可能出現的記憶衰減問題，使模型能夠理解整體揮拍過程中的結構性變化與微妙節奏。

此種架構不同於過去文獻中常見的單一 LSTM 或純 CNN 模型，能夠在保留時間資訊完整性的同時，強化模型對於動作模式的判別能力與細節捕捉能力，進而有效提升分類準確率與泛化能力。

在預測策略上，我們採用一種多樣本融合機制以提升輸出穩定性與解釋性。每筆測試資料會被切割成 27 段固定長度的子樣本，模型對每段分別進行獨立預測，並輸出對應的類別機率分佈。最終的預測結果非僅取單一最高機率輸出，而是透過 group-wise 機率平均與最大置信度重加權方式進行整合。此舉可有效消弭部分樣本中的預測波動，強化整體預測的一致性與魯棒性，並更具代表性地反映整段動作的特徵趨勢。

綜合上述設計，本研究的模型不僅在結構上融合不同時序理解模組，強化特徵擷取與關聯學習能力，在預測策略上也考量實務中動作表現的連貫性與變異性，整體大幅提升了系統的分類表現、穩定性與應用潛力。

## 肆、資料處理

### 1. 切分資料

原始資料為每位選手的多筆 .txt 檔案，每檔案記錄一段揮拍動作過程，包含來自六軸感測器的數據：加速度（ax, ay, az）與角速度（gx, gy, gz）。這些數據為連續時間序列，取樣率固定。為使模型能處理固定長度的輸入，我們將每段資料切割為 27 等長的子樣本，每段涵蓋相同的時間跨度與感測維度。針對部分資料長度不足 27 段的情況，我們於 `test_data_generate()` 函式中設計了自動補齊機制，將不足之處以最後一筆樣本重複補齊，確保每筆輸入資料皆為一致長度，有利於模型批次訓練與推論。

### 2. 特徵提取

每段子樣本資料會進行特徵工程處理，融合時域與頻域資訊，讓模型能兼顧動作在時間上的變化與其頻譜特性。透過 `FFT_data()` 與自製的 `FFT()` 函式，對合成訊號進行快速傅立葉轉換（FFT），提取其在頻域上的功率分佈與能量密度，反映動作的節奏特性與強度變化。除此之外，亦額外計算 spectral entropy 作為訊號複雜度指標，有助於模型分辨穩定與不穩定動作的區別。

最終，每段子樣本會轉換為一組固定維度的特徵向量，包含能量、頻率分佈、熵值等多重維度資訊，提供模型更全面的輸入基礎。

### 3. 分群與標籤對齊

在資料的分群部份我們是以 `player_id` 為依據使用 `train_test_split()` 將選手資料切為訓練集與驗證集，確保同一選手不會同時出現在兩者，防止資料洩漏。每筆 .csv 特徵資料對應回原始資料的 `unique_id`，並搭配 `train_info.csv` 中的標籤（gender、hold racket、play years、level）進行對齊，確保模型輸入與輸出對應正確。

## 伍、訓練方式

訓練流程由 `main()` 函式統籌：讀取 `train_info.csv`，並依據分類任務類型分別建立模型。對於 gender 與 hold hand 任務，使用 `CatBoostClassifier` 並以

`predict_proba()` 取得機率輸出；對於 `play years` 與 `level` 多類別任務，使用自建 CNN-BiLSTM-Transformer 模型，訓練過程記錄每輪 `loss`，最終儲存 `.pth` 權重。訓練後各模型預測結果會進行 `group-wise` 機率平均，並輸出至 `valid_prediction.csv`。若使用者選擇執行測試預測 (`predict_new_data()`)，則會自動載入訓練階段儲存之 `scaler` 與模型，並輸出 `test_prediction.csv`。整體流程模組化程度高，可重複訓練與推論。

## 陸、分析與結論

### 1. 結論

本專案所建構之分類系統成功整合 CatBoost 與 CNN-BiLSTM-Transformer 架構，有效應對桌球揮拍分類任務之多樣化目標。根據驗證集計算的 AUC 指標：`gender` 及 `hold` 任務皆達高於 0.95 的表現，`play years` 與 `level` 的平均 multi-class AUC 則亦穩定，證明深度模型能捕捉動作時間特徵的有效性。並且混合式架構結合傳統分類器與深度模型能有效對應不同任務特性，預測整合策略對結果影響顯著，是準確率穩定的關鍵。

### 2. 模型使用上的策略與想法

- CatBoost 為二元任務提供穩定表現，訓練快速、效果佳，適合小樣本應用
- 深度模型雖訓練時間長，但透過 CNN-LSTM-Transformer 混合可有效捕捉時序關聯與動作變化
- `group-wise` 預測可避免單一段異常導致整筆樣本錯誤，是一項關鍵技巧
- 適當的模型選擇與訓練參數設定。相較僅使用傳統統計特徵的 `baseline` 模型，本架構可提升整體分數逾 5%

### 3. 未來優化方向

- 引入 `attention` 可視化機制以提升模型可解釋性
- 利用增強學習自動探索最有效的切分方式與特徵選擇策略
- 將模型移植至低功耗裝置以利現場即時分析應用

- 將預處理的部分作加強，在特徵提取的部分可以將各特徵的分佈作平衡優化

## 柒、程式碼

Github：[https://github.com/tzuyu-TYC/AI\\_CUP.git](https://github.com/tzuyu-TYC/AI_CUP.git)

## 捌、使用的外部資源與參考文獻

- 機器學習入門必看！5 分鐘搞懂機器學習定義、類型＋應用 ...
- Day 20. 深度學習模型- Attention 和 Transformer - iT 邦幫忙