

# 基于深度学习的网球比赛动量分析与预测研究

张子诺, 南京大学信息管理学院, 231820015@smail.nju.edu.cn

**摘要**——本研究基于 2024 年温布尔登网球公开赛逐分数据, 采用深度学习模型识别并预测比赛中的“势头”变化。研究构建了包含得分差距、近期得分率、连胜次数等动量特征的时间序列数据集, 使用 LSTM 和 Transformer 进行对比实验。模型以历史 10 个点的特征序列预测下一分胜负, 采用 70%-15%-15%的数据划分, 使用 Adam 优化器训练, 并通过早停机制防止过拟合。实验采用准确率、F1 分数、AUC 等指标进行评估。结果表明, Transformer 在准确率和 F1 分数上表现更优, 能有效捕捉长期依赖和势头变化模式。研究将主观的“势头”概念量化为可计算指标, 为体育数据分析、智能赛事预测提供广阔支持。研究局限在于数据规模有限、未考虑外部因素等, 未来可通过数据增强、特征优化和模型改进进一步提升性能。

**关键词**——时间序列; 动量分析; LSTM; Transformer; 特征工程; 滑动窗口

## 1. 引言

### A. 研究背景

“势头”是体育比赛中影响比赛进程的重要因素, 它描述了一方选手在连续得分或关键分获胜后获得的心理与战术优势。在网球比赛中, 准确识别并预测势头变化对结果有重要意义。传统的体育数据分析多依赖专家经验与简单统计, 难以捕捉比赛的动态变化和连续事件之间的关联。近年来, 深度学习在时间序列预测中表现突出, 特别是长短期记忆网络和 Transformer 在序列建模中应用广泛。LSTM 能捕捉长期依赖, 适合建模比赛中的历史影响; Transformer 通过自注意力机制能关注全局依赖, 为体育数据分析提供了新的技术路径。本研究结合深度学习模型与逐分数据, 构建量化的“势头”指标体系, 并基于 LSTM 和 Transformer 模型进行对比实验, 探索深度学习在网球比赛动量分析与预测中的应用。

### B. 研究问题

如何基于逐分数据量化比赛中的“势头”变化? 能否使用深度学习模型预测下一分的胜负结果? 不同深度学习模型的预测性能如何?

### C. 研究目标

主要目标包括:

#### 1) 构建时间序列数据

将逐分数据转换为序列形式, 使用历史 10 个点的特征序列预测下一个点的结果。

#### 2) 量化动量特征

建立包含得分差距、近期得分率、连胜次数等指标的动量特征体系, 将“势头”概念量化。

#### 3) 模型训练与评估

分别训练 LSTM 和 Transformer 模型, 采用准确率、F1 分数、AUC 等指标评估性能。

#### 4) 模型对比分析

对比两种模型的预测性能, 分析各自特点。

#### 5) 应用价值探索

为体育数据分析、智能赛事预测和策略评估提供数据支持与方法参考。

## II. 数据来源与处理

### A. 数据来源

本研究使用 2024 年温布尔登网球公开赛的逐分比赛数据。数据集包含多场比赛的逐分记录, 每一条记录对应比赛中的一个得分点, 涵盖比赛过程中的关键信息。数据字段主要包括: 比赛 ID、选手姓名、经过时间等比赛基本信息; 盘数、局数、点数等比分信息; Ace 球、制胜分、非受迫性失误、双误等技术统计; 选手 1/2 跑动距离、发球速度等体能数据; 每分获胜者、每局获胜者、每盘获胜者等结果信息。数据按比赛 ID 和点数排序, 确保时间序列的连续性和正确性。

### B. 数据清洗

对缺失值进行处理, 数值型特征(如距离、速度)用 0 填充, 其他数值特征用中位数填充, 分类型特征(如发球宽度、发球深度)用 'Unknown' 填充。按比赛 ID 和点数排序, 确保时间序列连续性, 并识别处理异常数据点。对于关键列(match\_id、point\_victor、p1\_points\_won、p2\_points\_won), 删除包含缺失值的记录。

### C. 特征工程

为量化“势头”, 我构建了动量特征体系如下, 这些特征共同反映了比赛动态和选手状态。

#### 1) 基础比分特征

得分差距、盘数差距、局数差距。

#### 2) 动量特征

采用滑动窗口计算近期得分率, 窗口大小为 5; 统计连胜次数。

#### 3) 技术统计特征

Ace 球、制胜分、失误等累积统计, 发球效率、破发点转化率等。

- 4) 体能特征  
跑动距离、发球速度等。

#### D. 时间序列构建

将逐分数据转换为时间序列格式。使用前 10 个点的特征序列预测下一个点的结果，特征维度为 20+。标签为下一分的获胜者，转换为二分类，其中 0 表示选手 2 获胜，1 表示选手 1 获胜。数据集是按 70%、15%、15%划分为训练集、验证集和测试集。同时，为避免数据泄露，我在构建序列时仅使用历史数据预测时刻  $i$  的结果，不使用未来信息。最后，我对数值特征进行了标准化处理，以提升训练稳定性和收敛速度。经过上述预处理，原始数据转换为可用于深度学习模型训练的时间序列数据集。

### III. 得分预测模型设计与实现

#### A. 模型架构

本研究采用 LSTM 和 Transformer 两种深度学习模型进行对比实验，用于时间序列预测任务。

##### 1) LSTM 模型

采用 LSTM + Dropout + Dense 结构。LSTM 层用于捕捉长期依赖，设置隐藏层维度为 128、层数为 2，采用 Dropout (0.3) 降低过拟合，最后通过全连接层输出二分类结果。该模型适合处理时间序列，能捕捉比赛中的长期依赖关系。

##### 2) Transformer 模型

采用输入投影层、位置编码、Transformer 编码器和输出层的结构。模型维度设为 128，注意力头数为 8，编码器层数为 2，同样采用 Dropout (0.3) 正则化。自注意力机制可捕捉全局依赖，与 LSTM 形成互补对比。

#### B. 训练策略与实验环境

训练采用 Adam 优化器，学习率 0.001。损失函数为交叉熵，适合多类别分类任务。训练 50 轮，采用早停机制，当验证集损失不再下降时提前停止，防止过拟合。批次大小设为 32。同时，使用 WandB 记录训练过程指标，确保实验可复现。实验在 Python 3.8+ 环境下进行，采用 PyTorch 2.0+ 框架实现。使用 Hugging Face Datasets 管理数据集，便于数据共享与版本控制。

#### C. 模型评估

训练完成后，加载验证集上表现最好的模型权重，在测试集上评估模型性能，计算准确率、F1 分数、AUC、等指标。对 LSTM 和 Transformer 两个模型分别进行评估，对比各项性能指标。最后将测试结果保存为 JSON 文件，便于后续分析和报告撰写。

#### D. 可视化生成

##### 1) 训练曲线

绘制训练损失、验证损失、准确率、F1 分数、AUC 随训练轮数的变化曲线。

##### 2) 模型对比

绘制两个模型在各项指标上的对比表格及柱状图，直观展示性能差异。

##### 3) 预测趋势

绘制模型在测试集上的预测概率趋势图，展示预测概率与真实标签的对比，以及滑动窗口准确率趋势。

##### 4) 动量热力图

选择 3 场代表性比赛，绘制动量特征变化热力图，直观展示比赛中的势头变化。

### IV. 可视化结果分析

#### A. 训练曲线

下图 1 图 2 是 LSTM 和 Transformer 的训练曲线，采用早停 (patience=10)，监控验证损失。若连续 10 个 epoch 验证损失不下降，则提前停止，避免过拟合，并保存验证集上表现最好的模型。

##### 1) LSTM 模型

训练 12 个 epoch 后停止。训练损失从 0.67 降至 0.54，验证损失在 epoch 7 后由 0.67 升至 0.76，出现明显过拟合。验证准确率、F1 分数和 AUC 波动较大，模型表现不稳定。

##### 2) Transformer 模型

训练 14 个 epoch 后停止。训练损失收敛至 0.66-0.67，验证损失波动较大，存在轻微过拟合。验证准确率在 0.55-0.62 之间波动，F1 分数在 epoch 9 出现明显下降 (0.60 降至 0.35)，随后恢复。早停机制保存了验证集上的最佳模型权重。

##### 3) 对比总结

根据测试结果，LSTM 的准确率 (0.6199) 略高于 Transformer (0.6132)，但 Transformer 的 F1 分数 (0.6087) 略高于 LSTM (0.5939)。两者 AUC 相近 (LSTM 0.6507, Transformer 0.6427)。早停机制有效防止了过拟合的进一步恶化，并保存了最优模型权重，但两模型的训练稳定性均有待提升。

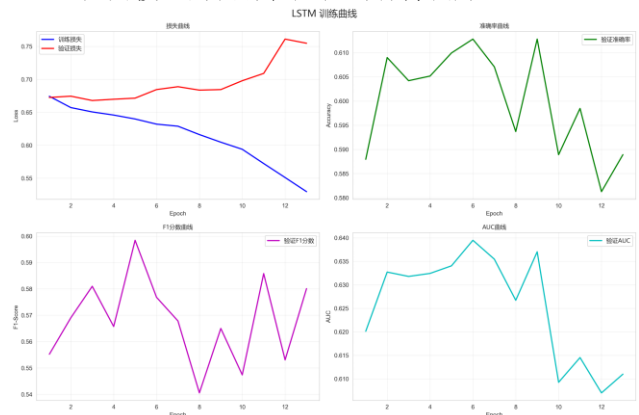


图 1 training\_curves\_LSTM

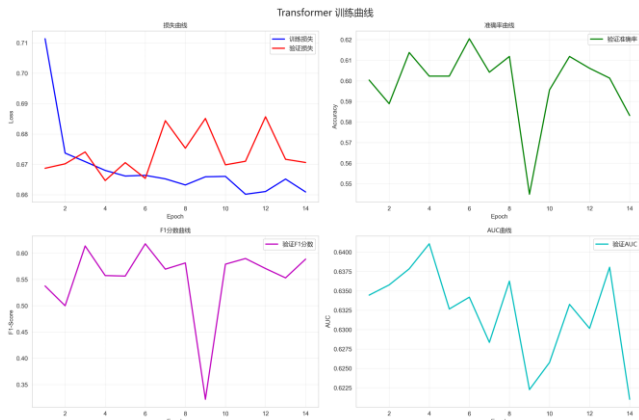


图2 training\_curves\_Transformer

## B. 模型对比

基于训练曲线图，我得到了模型对比表格和柱状图如下所示。

可以发现，Transformer 的 F1 分数更优，自注意力机制可捕捉全局依赖和跨时间步关系，在处理复杂模式和多特征交互上具有潜力，有利于捕捉比赛中的动态变化。虽然训练过程波动略大，但通过早停机制保存了最佳模型权重，保证了性能。LSTM 模型在准确率和 AUC 上略优，训练过程相对稳定，但在 F1 分数上略低，且主要依赖短期记忆，在处理长程依赖和复杂特征交互时可能受限。

综合评估来看，虽然两个模型性能接近，但考虑到 F1 分数在该预测任务中的重要性，以及 Transformer 在特征交互和动态模式捕捉上的潜力，选择 Transformer 模型作为最终模型。虽然训练稳定性需要改进，但自注意力机制带来的全局依赖能力使其更适合捕捉比赛中的复杂动量变化。未来可通过超参数调整、正则化优化或模型集成进一步提升性能和稳定性。

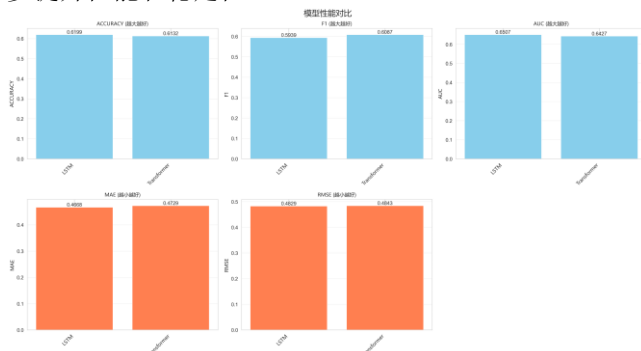


图3 model\_comparison

模型	LOSS	ACC	F1 分数	AUC
LSTM	0.659	0.619	0.594	0.651
Transformer	0.662	0.613	0.609	0.643

表1 model\_comparison

## C. 预测趋势

通过将得到的最优 Transformer 模型用于测试集，可以得到如下的预测趋势分析图。包括预测概率趋势和预测准确率趋势。

### 1) 预测概率趋势分析

预测概率趋势图展示了模型在测试集上的预测概率与真实标签的对比。图上显示，模型预测概率（蓝色线）在 0.0-1.0 之间波动，总体与真实标签（红色柱状图）的变化趋势一致。多数情况下，当真实标签为 1 时，预测概率多在 0.5 以上；当真实标签为 0 时，预测概率多在 0.5 以下，表明模型能够有效区分两类样本。

### 2) 滑动窗口准确率趋势分析

下图展示了采用滑动窗口（窗口大小=50）计算的预测准确率随时间的变化趋势。初始阶段（前约 50 个样本）由于窗口尚未填满，准确率波动较大；随着窗口填充，准确率逐渐趋于稳定。整体准确率在约 0.58-0.75 之间波动，反映了模型在不同样本段的表现差异。在样本序号约 100 和 700 附近出现峰值，说明模型在这些区间表现较好；而在其他时段准确率有所下降，可能受数据分布变化或模型对不同样本特征敏感度差异影响。

### 3) 综合评估

结合上下两图，模型在测试集上的整体表现稳定。预测概率能够较好地反映真实标签的分布，准确率在大部分时间段维持在合理水平。模型在不同样本段的性能波动表明存在提升空间，可通过进一步优化特征工程或模型架构来增强鲁棒性。滑动窗口准确率的波动也提示，模型对某些样本特征或数据分布变化的适应能力有限。未来可考虑引入自适应机制或集成方法，提升在不同数据段的表现一致性。



图4 prediction\_trend

## D. 动量热力图

下图 5 动量热力图展示了比赛 2023-wimbledon-1301 中 7 个动量特征随时间的变化，绿色表示选手 1 领先，红色表示选手 2 领先，黄色表示比分接近。这种可视化方法直观展示了比赛的动态特征，验证了动量特征体系的有效性。

### 1) 得分差距

时间步 60-120 出现深红色，选手 2 领先；时间步 180 到结束转为深绿色，选手 1 反超并保持优势，显示明显的势头转换。

## 2) 盘数差距

主要为橙色、黄色和浅绿色，选手 1 在盘数上保持优势或接近持平。

## 3) 局数差距

与得分差距类似，但强度较弱。前期选手 2 短暂领先，后期选手 1 取得优势。

## 4) 近期得分率

p1\_recent\_win\_rate 和 p2\_recent\_win\_rate 主要为橙色和黄色，波动较小，双方近期得分率接近。

## 5) 连胜次数

p1\_streak 和 p2\_streak 呈现交替模式，未出现持续极端值。p2\_streak 在比赛末期出现浅绿色段，说明选手 2 在后期有连胜。

## 6) 势头转换识别

比赛分为三个阶段。第一阶段（1-120 步）选手 2 占优；第二阶段（120-180 步）为过渡期；第三阶段（180 步-结束）选手 1 反超并保持优势。

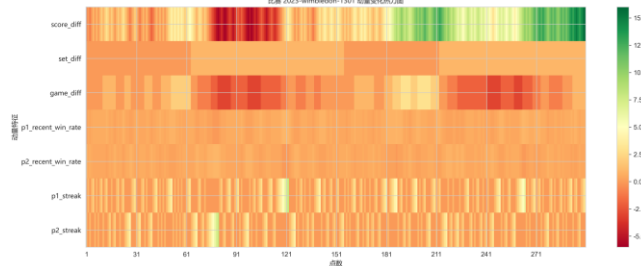


图 5 momentum\_heatmap\_1

## V. 研究结论

基于 2024 年温布尔登网球公开赛的逐分数据，我采用深度学习模型进行网球比赛动量分析与预测。通过构建动量特征体系，将“势头”这一主观概念量化为可计算的指标。不仅揭示了网球比赛中“势头”变化的动态规律，也为体育数据分析和智能赛事预测提供了可操作的工具与决策支持框架。通过系统对比 LSTM 和 Transformer 两种模型，为模型选择提供了参考依据；通过多种可视化方法，直观展示了模型能和比赛动态特征。

本研究验证了深度学习模型在体育数据分析中的有效性，为后续研究奠定了基础，也为类似体育项目的动量分析与预测提供了可借鉴的方法与实践经验。

## VI. 不足与展望

### A. 不足

尽管本研究在网球比赛动量分析与预测方面取得了一定成果，但仍存在若干不足。首先，数据来源主要依赖 2024 年温布尔登单一年份、单一赛事的数据，时间跨度有限，未能捕捉不同年份、不同赛事和不同场地类型的影响，同时外部因素如选手历史数据、心理状态、比赛环境等未纳入模型，可能对预测精度和泛化能力产生一定影响。其次，虽然通过构建动量特征体系成功量化了“势头”这一主

观概念，但对突发性事件和极端比赛情况的反应仍有限，模型在异常情况下的鲁棒性有待提升，训练过程中出现的波动性也表明模型对数据分布和训练动态较为敏感。

### B. 展望

在未来研究中，可以从多个方向进一步拓展。如引入更多维度的多源数据，包括不同年份、不同赛事、不同场地类型的数据，以及选手历史数据、比赛环境因素等，以增强模型对不同比赛情境的识别能力和泛化能力。其次，可探索更复杂的深度学习架构或图神经网络方法，结合时空依赖建模和特征交互，更精准地刻画比赛中的动态变化规律和选手之间的交互关系。同时加强模型的可解释性研究，通过注意力机制可视化和特征重要性分析，为模型预测提供更可靠的解释，为体育数据分析和智能赛事预测提供更有效的数据支撑和决策支持。

## 附录

1. 全部代码已上传至 GitHub，链接：  
<https://github.com/zzn20230810/Final-assignment.git>
2. 数据集可通过 Hugging Face 平台获取，链接：  
[https://huggingface.co/datasets/saxasxa/2024\\_Wimbledon\\_featured\\_matches](https://huggingface.co/datasets/saxasxa/2024_Wimbledon_featured_matches)
3. 所有实验记录通过 WandB 平台进行管理，链接：  
<https://wandb.ai/zzn20230810-/tennis-momentum?nw=nwuserzzn20230810>
4. 可视化成果见 visualizations 文件夹。
5. 程序使用与环境配置见 README.md 和 requirments.txt 文件。