Modeling and Simulation 建模与仿真, 2024, 13(3), 2259-2267
Published Online May 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.133207

Hans汉斯

基于LASSO回归的BP-LSTM模型对网球比赛势头的研究

刘甜甜，陈丽娜

上海理工大学健康科学与工程学院，上海

收稿日期：2024年4月15日；录用日期：2024年5月13日；发布日期：2024年5月20日

摘　要

在各大体育赛事中，球员在比赛中的优势表现往往对比赛的结果至关重要。为了定量化分析势头对各体育赛事的影响，我们将势头概念量化，引入势头得分这一概念，并随机抽取五场网球比赛进行分析，将破发成功率，二发成功率，一发成功率，一发得分率作为影响势头分的指标，通过LASSO回归建立势头得分与各指标的量化关系，考虑到多指标的输入输出，我们引入BP神经网络建立起多指标对输出的关系，并将LASSO回归得到的具体的势头分作为BP神经网络的输出，将BP神经网络残差输出借助LSTM预测，并将二者相加得到较为精准的势头得分，正确率高达95%。结果表明各球员在比赛中的势头在一定程度上影响比赛的结果。

关键词

势头，LASSO，BP神经网络，LSTM

Research on the Momentum of Tennis Matches Using the BP-LSTM Model Based on LASSO Regression

Tiantian Liu, Lina Chen

School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 15th, 2024; accepted: May. 13th, 2024; published: May. 20th, 2024

Abstract

The dominant performance of players in major sports events often plays a crucial role in deter-

文章引用: 刘甜甜, 陈丽娜. 基于 LASSO 回归的 BP-LSTM 模型对网球比赛势头的研究[J]. 建模与仿真, 2024, 13(3): 2259-2267. DOI: 10.12677/mos.2024.133207

mining the outcome of the game. To quantitatively analyze the impact of momentum on various sports events, we have defined and measured the concept of momentum score. Five randomly selected tennis matches were analyzed, considering indicators such as break success rate, second serve success rate, first serve success rate, and overall first serve success rate that influence the momentum score. Through LASSO regression analysis, we established a quantitative relationship between each indicator and the momentum score. Taking into account multiple input-output indicators, we introduced a BP neural network to establish the relationship between these indicators and output. The specific momentum score obtained from LASSO regression was used as an output for BP neural network prediction while LSTM was employed to predict residual outputs from BP neural network. By combining these two approaches, we achieved a relatively accurate prediction of momentum scores with an accuracy rate of 95%. These results demonstrate that individual player's momentum during a match has some degree of influence on its final outcome.

Keywords

Momentum, LASSO, BP Neural Network, LSTM

1. 引言

网球作为一项全球性的竞技运动一直以来广泛受到人们的关注。随着时代的变化以及数据科学的发展，以往通过人工观察和分析的方法难以准确评估比赛的结果，因此利用数据分析技术对网球比赛进行客观、全面、深入的分析已经成为了一种必要的趋势和发展方向[1]。本研究将球员在比赛中抽象的势头量化为势头得分，得分高者取得本场比赛的胜利。通过 LASSO 回归建立势头得分与影响势头得分的各指标之间的关系，根据回归结果可以对参加比赛的球员提供一些实质性的建议；同时 LASSO 回归得到的具体的势头分作为 BP 神经网络的输出，建立起多指标的输出关系，用 LSTM 对 BP 神经网络输出残差进行预测，得到较为精准的势头得分。通过势头分的高低判断球员比赛的状态，便于教练在赛场上及时做出决策，以此提升选手的表现，取得更优异的成绩。

2. 数据预处理

通过澳大利亚网球公开赛官网搜集球员比赛的指标，查阅相关文献后选定指标如下：ACE、破发率、非受迫性失误、一发成功率、一发得分率、二发成功率和双误，共七个变量，构建出如图 1 的三级指标体系。

为验证各指标之间的关联性，我们选取了 2023 年温布尔登男单决赛中五组比赛数据，对各指标进行斯皮尔曼相关系数的相关性检验，我们得到如图 2 所示的热图。

从图 2 中发现，这几个指标的相关性都很强，故将这七个指标作为衡量球员表现的重要因子。

介于三级指标体系中各指标大多数为定性变量，为方便后续计算与模型的建立，我们对部分指标进行量化，我们定义：

在每盘网球比赛中：

$$\text{一发成功率} = \text{一发的个数/发球总数}$$

$$\text{一发得分率} = \text{一发得分数/一发个数}$$

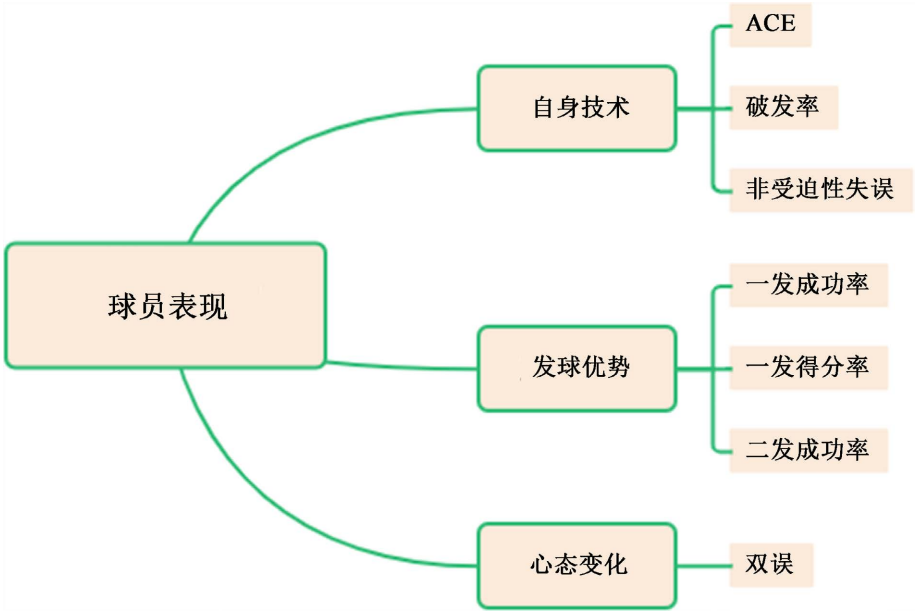

Figure 1. Three-level index system that affects player performance
图 1. 影响球员表现的三级指标体系

	ACE	Double fault	Unforced errors	First serve success rate	First serve scoring rate	Second serve success rate	Break rate
ACE	1.000	0.354	0.161	0.554	0.67	0.634	-0.561
Double fault	0.354	1.000	0.609	0.449	0.481	0.671	-0.528
Unforced errors	0.161	0.609	1.000	0.398	0.289	0.452	-0.481
First serve success rate	0.554	0.449	0.398	1.000	0.86	0.569	-0.831
First serve scoring rate	0.67	0.481	0.289	0.86	1.000	0.796	-0.832
Second serve success rate	0.634	0.671	0.452	0.569	0.796	1.000	-0.738
Break rate	-0.561	-0.528	-0.481	-0.831	-0.832	-0.738	1.000

Figure 2. Heat map of correlation of indicators at all levels
图 2. 各级指标相关性热图

二发得分率 = 二发得分数/(发球总数 − 一发个数)

破发率 = 我方得分数/对方发球总数

3. LASSO 回归势头得分的建立

3.1. 提出势头得分概念

网球比赛中各球员的势头是难以预测的，波动性较大，且势头本身是一个比较抽象化的概念，随着每场比赛时间的推移，应用系统响应时间的变化规律复杂多变[2]。为进一步建立数学模型，我们将势头概念量化，引入势头得分这一概念。我们将各球员表现的势头分作为衡量比赛中各球员的优势转向。依据 P1 − P2 的基本准则，若 P1 在本局比赛中势头分高，即 P1 − P2 > 0，则优势转向 P1，反之亦然。

3.2. 建立 LASSO 回归

为提高势头分定义的准确性，我们将破发率、二发得分率、一发成功率、一发得分率、ACE 和双误这六个相关系数较高的指标作为自变量，势头得分作为因变量，并对二者进行 LASSO 回归，引入惩罚项识别模型中相对不重要的因子。

LASSO 回归表达式如下：

$$MS = 83.23BR + 47.42FSS\_success + 29.43FSS\_score + 20.72SSS + 2.23ACE - 2.32DF - 0.249$$

其中 MS 代表势头分，BR 代表破发率，FSS_success 代表一发成功率，FSS_score 代表一发得分率，SSS 代表二发得分率，ACE 代表通过使对手无法回发球而发球得分的个数，DF 代表双误。

通过 LASSO 回归表达式我们可以计算出各选手在比赛任意一局的势头得分，这有助于评估选手当前的比赛状态和势头优劣。教练可以根据这些预测结果，及时调整战术和策略，以应对不同的比赛情况。例如，当发现对方选手势头得分较高时，可以采取防守策略；当自己选手势头得分较高时，可以采取进攻策略。

4. BP-LSTM 时间序列预测模型的建立

为了进一步预测网球比赛的比赛结果，我们引入了 BP-LSTM 时间序列模型，将势头得分的高低作为比赛输赢的依据，结合上述六个指标进行结果预测，可以得到更具有解释性和预测准确性的模型。在这个模型中，BP 神经网络负责提取非线性特征，LSTM 负责处理时间序列数据，结合起来可以更好地捕捉比赛中势头得分的变化和趋势，从而更准确地预测比赛结果。

Step 1：

考虑到多指标输入和双方势头分输出的多样性，引入 BP 神经网络算法，BP 神经网络的数学表达式如下：

$$\hat{y} = g_2\left(V^{\mathrm{T}} g_1\left(W^{\mathrm{T}} X + b_1\right) + b_2\right) \tag{1}$$

在我们建立的神经网络模型中，输入层为三级指标体系中的 ACE，DF，FFS_success，FFS_score，SSS，BR，UE 等 7 个指标，输出层为 P1 和 P2 在 LASSO 回归处理下对势头得分的量化。

其中输入层到隐藏层的权重值为 W，偏置项为 b_1，激活函数为 g_1；隐藏层到输入层的权重值为 V，偏置项为 b_2，激活函数为 g_2。

输入层到隐藏层：

$$Net_1 = w^{\mathrm{T}} x + b_1, h = g_1\left(Net_1\right) \tag{2}$$

隐藏层到输出层：

$$Net_2 = v^{\mathrm{T}} h + b_2,\ \hat{y} = g_2\left(Net_2\right) \tag{3}$$

损失函数：

$$E = \frac{1}{n}\sum_{i=1}^{l}\left(y_i - \hat{y}_i\right)^2 \tag{4}$$

然而在真实的训练过程中，BP 神经网络算法中隐藏层中的神经元个数是难以确定的，我们一般通过经验获取得到，但是隐藏层神经元的个数对神经网络的收敛性，准确性和收敛速度有着很大的影响，我们只能通过不断调整隐藏层神经元的个数和训练样本的数量来确定最后迭代的次数。

我们将数据导入到 MATLAB 中，通过不断调整迭代的系数，最后将隐藏层的神经网络参数定为 20 个，具体的神经网络参数如图 3。

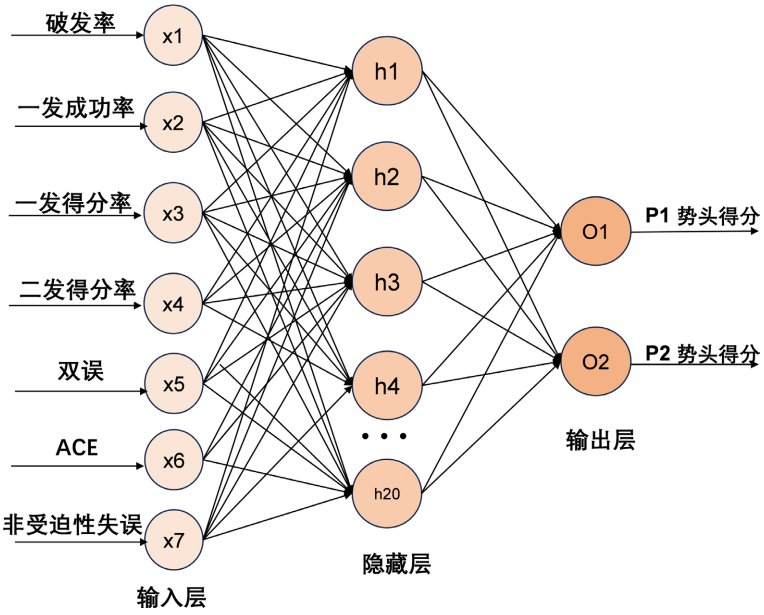

Figure 3. Neural network model diagram
图 3. 神经网络模型图

为了提高最后预测结果的准确性，我们使用 70%的数据作为训练集，15%的数据作为验证集，15%的数据作为测试集，随着迭代次数的不断增加，误差不断地减小，当迭代次数达到 51 次时，训练组的均方误差(MSE)达到最小为 0.3977，训练停止。

从训练结果可以看出测试集的训练结果达到 0.99，如图 4，因此我们可以认为 BP 神经网络可以很好的预测各球员的势头得分，记为 L_t。

Step 2：

为了提高模型的准确性，我们将使用 LASSO 回归得到的势头分与 BP 神经网络预测的势头分相减得到残差，作为输入在 LSTM 模型如图 5 中训练，从而得到势头分的预测残差。

最终得到非线性部分残差的预测值，其数学表达式如下：

$$N_t = f\left(e_{t-1}, e_{t-2}, \cdots, e_{t-m}\right) + \varepsilon_t \tag{5}$$

其中 e_t 为残差。

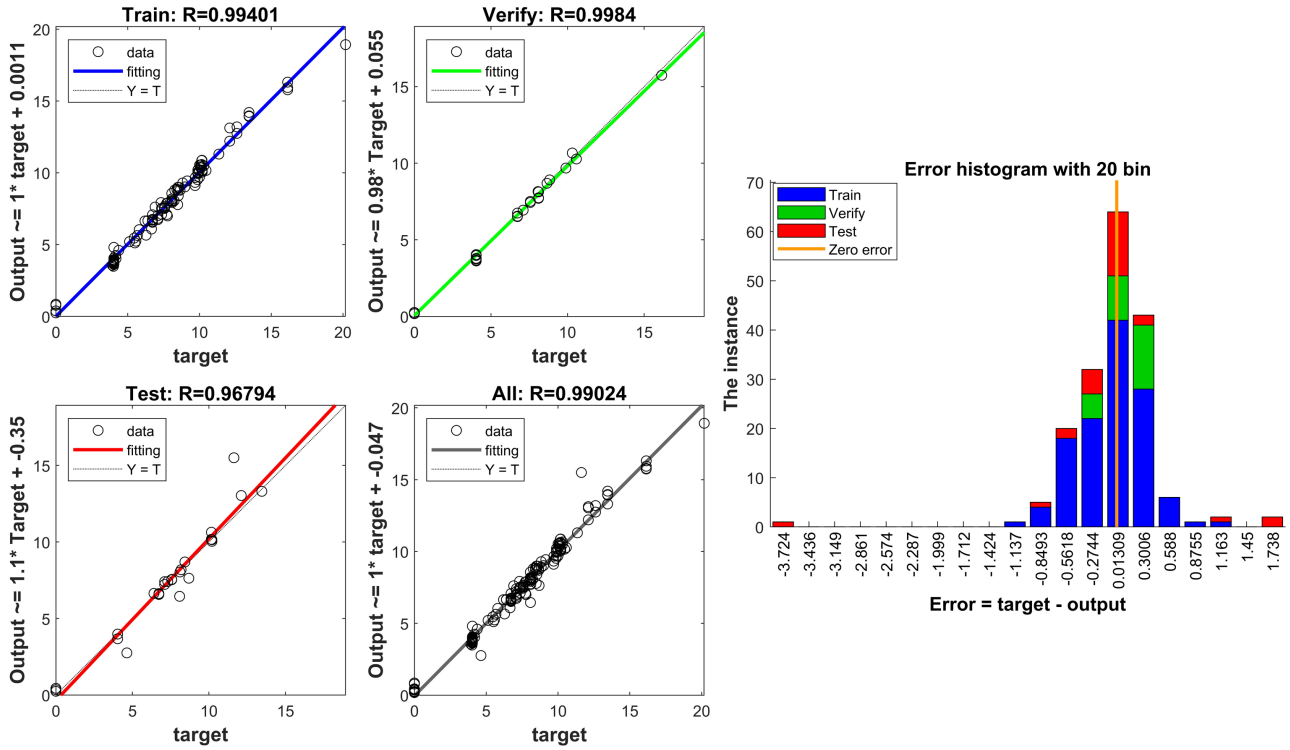

Figure 4. Data training results
图 4. 数据训练结果

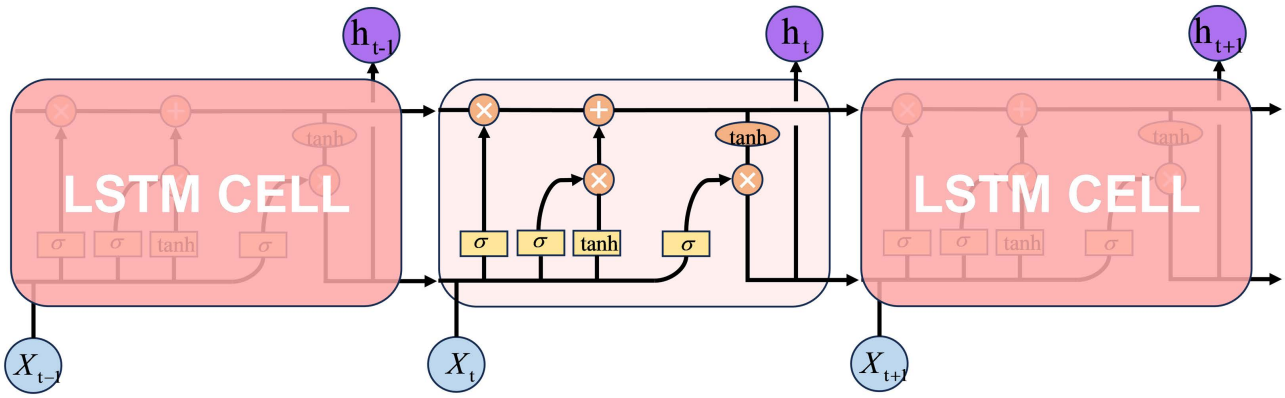

Figure 5. LSTM model structure
图 5. LSTM 模型结构

将 L_t 和 N_t 相加可以得到最终预测结果：

$$y = L_t + N_t \tag{6}$$

从图 6 中可以近似看出，BP 混合 LSTM 神经网络组合模型加强了神经网络对时序数据的反应能力，使得预测结果与真实结果的拟合度达到更高的水准[3]。表 1 为单独用 BP 神经网络进行预测的结果与 BP-LSTM 组合模型对比赛输赢的预测结果比较，BP 预测结果的准确率为 85%，BP-LSTM 的准确率为 90%如表 1，由此可见 BP 混合 LSTM 神经网络组合模型的正确率更高。

Figure 6. Comparison of BP (left) and BP-LSTM (right) with prediction results
图 6. BP(左)与 BP-LSTM(右)与预测结果比较

Table 1. Predicting match results with BP vs. predicting match results with BP-LSTM combination
表 1. BP 预测比赛结果与 BP-LSTM 组合预测比赛结果

BP				BP-LSTM			
P1	P2	预测输赢	实际输赢	P1	P2	预测输赢	实际输赢
10.1	3.67	1	1	10.37	4.04	1	1
0.25	10.07	0	0	0.50	10.00	0	0
8.19	7.44	1	1	7.55	7.57	0	1
5.81	6.60	0	0	5.62	8.10	0	0
8.57	8.45	1	0	9.27	8.41	1	0
7.22	8.05	0	0	7.25	8.11	0	0
10.1	4.02	1	1	10.04	4.04	1	1
15.5	18.84	0	1	12.93	3.43	1	1
3.81	16.15	0	0	3.04	16.15	0	0
6.11	7.02	0	0	6.23	7.55	0	0
3.01	13.81	0	0	0.13	13.46	0	0
3.56	9.52	0	0	3.51	10.16	0	0

续表

7.93	7.98	0	1	8.20	7.57	1	1
−0.09	10.01	0	0	−0.17	10.19	0	0
8.59	6.53	1	1	9.03	6.73	1	1
16.9	2.72	1	1	15.94	4.00	1	1
8.52	4.92	1	1	8.74	4.04	1	1
3.33	10.63	0	0	3.07	10.16	0	0
7.36	6.93	1	1	7.08	6.73	1	1
11.5	3.65	1	1	11.79	6.09	1	1

5. 给进入比赛的球员的建议

网球比赛的外在表现形式是战术，而内在则为策略[4]，因此需要球员在进入比赛前就了解何为比赛的制胜因素，以便在比赛过程中及时调整策略，提高比赛的胜率。

为了能更进一步直观显示出哪些指标对势头得分的影响最大，我们另选了本场比赛其余选手的比赛数据，建立起了具有泛化性的 LASSO 回归模型，各指标的回归系数相当于在每场比赛中影响各球员势头的重要程度。

$$MS = 25.05BR + 16.62FSS\_score + 14.75FSS\_success + 2.21SSS - 0.57DF + 1.26 \quad (7)$$

从该回归系数我们可以推断势头分与破发率、一发得分率、二发成功率成正相关，其中破发率对势头分的影响最大，一发得分率次之，势头分与双误成负相关。

Table 2. Momentum Scores for Players in the 20th and 25th games
表 2. 第 20 局和第 25 局各球员比赛势头分

	BR	FSS_score	FSS_success	SSS	DF	MS
The 20th game						
Jiri Lehecka (P1)	0.156	0	0	0	0	5.1649
Tommy Paul (P2)	0	0.188	9.171	0.182	0	6.0630
The 25th game						
Jiri Lehecka (P1)	0	0166	0.235	0.094	0	7.6127
Tommy Paul (P2)	0.130	0	0	0	0	3.2667

(1) 结果

从表 2 中我们可以得到在第 20 局比赛中，球员 P2 的势头分为 0.22681，球员 P1 的势头分为 0.2117，球员 P2 以略微的优势赢得了球员 P1，在第 25 局比赛中，球员 P1 的势头分为 0.2532，球员 P2 的势头分为 0.148，球员篇以将近 0.1 分的优势赢了球员 P1。

(2) 分析

通过分析数据我们发现当球员作为发球方时，赢得的几率更大，但并不是作为发球方就一定能够获胜，同时要提高发球的质量，降低双误的几率；而对于接发球方，并不是坐以待毙，若是能掌握接发球技巧，乘胜追击，同样能提高获胜的几率。在第 20 局中，P2 即使作为发球方，但是发球质量不够高，并且有一次双误，给了对方球员 P1 反追的机会，因此 P1 较于 P2 只有略微优势。在第 25 局中，P1 作为

发球方，提高了发球质量，P2 虽然成功破发对方一个球，但是介于对方发球质量相对较高，所以 P1 优势遥遥领先于 P2。

(3) 建议

通过查找资料，以及在各大比赛中网球球员所用策略，我们建议：

a) 提高发球质量，运用发球优势凸显进攻意识

全面提高发球质量，一发时以平击发球为主，突出速度优势，二发时提高旋转运用能力，注意落点的变化[3]，在此基础上提高发球稳定性。

b) 提高接发球质量，缩短攻守转换时间

改变“打回去就好”的战术指导思想，争取利用一切机会主动发起进攻。

6. 结论与展望

随着数字化的发展，利用数字分析技术对网球比赛进行全面分析已经成为发展的趋势。本研究提出利用 LASSSO 回归计算出球员在比赛中的势头得分，经验证，本方法具有一定的可靠性，再通过 BP-LSTM 混合模型对比赛结果进行预测，得到了较高的准确率，说明通过比赛数据对比赛结果进行预测的可行性。同时，我们选取了其它几组比赛的数据进行 LASSO 回归，基于 LASSO 回归结果我们分析给出了各球员在进入比赛时的策略建议。

本文提出的模型能够量化球员在比赛过程中的势头，较为准确地预测出本场网球比赛的赛果，通过 LASSO 回归自动进行特征选择，利用惩罚项将模型中不重要的特征系数置零，从而得到影响势头得分的重要指标模型。使用 BP-LSTM 预测势头得分，通过 BP 神经网络引入更多相关特征，提供更全面的信息，从而提高预测模型的准确性和稳定性。通过比较和综合不同模型的结果来降低预测的不确定性，提高了结果预测的可信度。此外，深入地评估比赛不同时刻球员所表现出的势头得分，可以帮助教练及时在比赛中制定合理的战术决策，增加取得本场比赛胜利可能性，赛后还能通过比赛数据分析最相关的绩效指标，为运动员制定个性化训练。

参考文献

[1] 王秋实, 加林·加恒努尔, 丁澜, 等. 基于职业网球开源信息大数据分析的研究进展[C]//中国体育科学学会. 第十三届全国体育科学大会论文摘要集——墙报交流(运动训练学分会) (二). 2023: 3. https://doi.org/10.26914/c.cnkihy.2023.074111

[2] Ni, X. and Chen, Y. (2024) Establishment of a Drug Procurement Decision Prediction Model Based on ARIMA Model and LSTM Neural Network. *Heilongjiang Science*, **15**, 76-78.

[3] Yao, Y., Hong, R. and Liu, Q. (2023) Research on Carbon Price Prediction Based on BP-LSTM Hybrid Neural Network. *Environmental Science and Management*, **48**, 71-76.

[4] 王伟, 周曙. 职业网球竞赛博弈特征研析[C]//湖北省体育科学学会. 第一届湖北省体育科学大会论文集(第一册). 黄石: 湖北师范大学, 2023: 3. https://doi.org/10.26914/c.cnkihy.2023.078040