

基于LSTM模型的草原土壤状态预测

吕园园, 赵大胜, 金志尧

上海理工大学机械工程学院, 上海

收稿日期: 2024年1月5日; 录用日期: 2024年3月8日; 发布日期: 2024年3月19日

摘要

中国是一个资源大国, 拥有丰富的草地。这些草原生态系统是保持我国生态平衡的关键防线, 同时为经济发展提供稳固的基础。然而, 随着畜牧业的迅速发展, 草地退化问题日益突出, 甚至在一些地区呈现沙漠化趋势。面对这一情况, 迫切需要提供科学合理的草地管理方式。因此, 对土壤状态的准确预测对于草原的可持续保护和合理开发至关重要。本文通过对历年统计数据进行分析, 首先通过Softmax逻辑回归模型, 结合处理后的数据, 得到六种土壤状态与不同放牧策略的四分类模型, 建立了不同放牧策略对草原土壤状态影响的数学模型。然后使用LSTM模型并基于多年份同放牧强度和放牧小区的土壤不同状态下的数据来进行训练, 预测得到2023年的同条件下的土壤在不同状态下的数据。再利用沙漠化程度指数预测模型和数据来确定不同放牧强度下监测点的沙漠化程度数值。最后使用有机量、含水量、叶面积指数等指标代替用于衡量土壤状态的指标, 有土壤肥力变化、土壤湿度、植被覆盖等, 综合2014~2022年的土壤数据, 使用LSTM模型预测得到2024年同月的土壤状态数据。

关键词

Softmax逻辑回归模型, LSTM模型, 衡量指标, 土壤状态

Grassland Soil Condition Prediction Based on LSTM Model

Yuanyuan Lyu, Dasheng Zhao, Zhiyao Jin

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

Received: Jan. 5th, 2024; accepted: Mar. 8th, 2024; published: Mar. 19th, 2024

Abstract

China is a resource-rich country with abundant grassland areas. These grassland ecosystems serve as crucial barriers to maintaining the ecological balance in the country while providing a stable

foundation for economic development. However, with the rapid expansion of livestock farming, the issue of grassland degradation has become increasingly prominent, even manifesting desertification trends in certain regions. Faced with this situation, there is an urgent need to establish scientifically sound grassland management approaches. Therefore, accurate prediction of soil conditions is crucial for the sustainable protection and rational development of grasslands. This paper analyzes historical statistical data and employs a Softmax logistic regression model to establish a four-class model for the impact of different grazing strategies on six soil conditions. This model is based on processed data, providing insights into the influence of various grazing strategies on grassland soil conditions. Subsequently, an LSTM model is utilized and trained on data from multiple years with consistent grazing intensities and different soil conditions in grazing zones, to predict soil conditions in 2023 under the same conditions. Utilize a Desertification Severity Index Prediction Model and data to determine the degree of desertification at monitoring points under different grazing intensities. Finally, employing indicators such as organic matter, moisture content, and leaf area index as substitutes for metrics measuring soil conditions like soil fertility changes, soil moisture, and vegetation cover, the LSTM model, incorporating comprehensive soil data from 2014 to 2022, predicts soil condition data for the same months in 2024.

Keywords

Softmax Logistic Regression Model, LSTM Model, Evaluation Metrics, Soil Condition

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

草原生态系统作为陆地生态系统的重要组成部分之一，植被类型分布广泛，是我国生态安全和食品安全的关键支柱。放牧优化问题是防止草原沙漠化的关键。过度放牧由于牲畜密度过大可能导致土壤裸露面积增加，对土地内水分相对运动产生不利影响，使土壤盐碱度程度增加，最终引发土壤沙漠化。相反，适度的放牧有助于提升土壤的状态，促进草原生态系统的可持续发展。放牧优化问题的研究也为国家、政府制定放牧政策和草原管理决策提供科学的依据[1] [2]。因此，预测不同放牧策略(放牧方式和放牧强度)对草原土壤性质的影响是非常有必要的。本文基于土壤有机碳数据、无机碳、全 N 以及土壤 C/N 比等值等信息，通过衡量土地状态的指标，土壤肥力变化、土壤湿度、植被覆盖等评估了土壤各性质对环境影响的响应程度。利用统计年鉴 2014~2022 年的数据，建立模型预测 2023 年在不同放牧策略(放牧方式和放牧强度)的情况下，使用 LSTM 法进行学习并对 2024 年九月的土地状态进行预测。

2. 模型建立与求解

本文用到的符号及其含义如表 1 所示。

2.1. Softmax 模型

针对问题中建立不同放牧策略(放牧方式和放牧强度)对锡林郭勒草原土壤化学性质影响的数学模型，首先需要通过逻辑回归方法来判断不同放牧策略对锡林郭勒草原土壤化学性质变化的发生概率大小，进一步通过 Softmax 模型获得四种土壤化学性质与不同放牧策略的四分类模型的权重值以及相应分类模型的常数值。

Table 1. Symbol description

表 1. 符号说明

符号	含义
RMSE	均方根误差
MAE	平均绝对误差损失
MAPE	平均绝对百分比误差
LSTM	长短期记忆预测模型
Q_i	因子强度
W_{ci}	权重系数

2.1.1. Softmax 建模

Softmax 模型事实上是逻辑回归(Logistics)模型的推广, Softmax 总目标是多分类, 也是在多元线性回归基础上把结果缩放到 0 到 1 之间。然后依据属于哪一类的概率最高, 选择该类, 其中线性回归函数的数学表达式:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \theta^T x \quad (1)$$

其中 x_i 是自变量, y 是因变量, y 的值域为 $(-\infty, \infty)$, θ_i 是常数项, $\theta_i (i=1, 2, \dots, n)$ 是待求系数, 不同的权重 θ_i 反映了自变量对因变量不同的贡献程度。

Softmax 模型:

$$\hat{y}^i = h_{\theta}(x^i) = \begin{bmatrix} P(y^i = 1 | x^i, q) \\ P(y^i = 2 | x^i, q) \\ \vdots \\ \vdots \\ \vdots \\ P(y^i = k | x^i, q) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{-\theta_j^T x^i}} \begin{bmatrix} e^{-\theta_1^T x^i} \\ e^{-\theta_2^T x^i} \\ \vdots \\ \vdots \\ \vdots \\ e^{-\theta_k^T x^i} \end{bmatrix} \quad (2)$$

\hat{y} 值: 概率含义(\hat{y} 越大说明该例概率越大)。

Softmax 回归模型如下图 1 所示。

2.1.2. 模型结果

首先需要通过逻辑回归方法来判断不同放牧策略对草原土壤化学性质变化的发生概率大小, 如表 2 所示。表中数据的出重度放牧“3”准确率最大, 中度放牧“2”准确率最低。

通过求解 Softmax 模型获得不同放牧策略对锡林郭勒草原土壤化学性质变化的概率权重, 即四分类模型系数如下:

$$\begin{bmatrix} 0.18749361 & -0.08152382 & 0.18908695 & -1.7968938 & 0.04560131 & 0.12562905 \\ -0.01772552 & 0.00408382 & 0.05685388 & 0.19276374 & 0.00868464 & 0.32389637 \\ 0.01157313 & 0.01591795 & -0.1700262 & -0.52627501 & -0.00860049 & -0.32719223 \\ -0.18134122 & 0.06152205 & -0.07293821 & 2.13040507 & -0.04568546 & -0.12233319 \end{bmatrix}$$

其中每个分类模型所含常数项如下：

$$[-6.24012973 \quad -2.86237095 \quad 5.54929601 \quad -0.60797045]$$

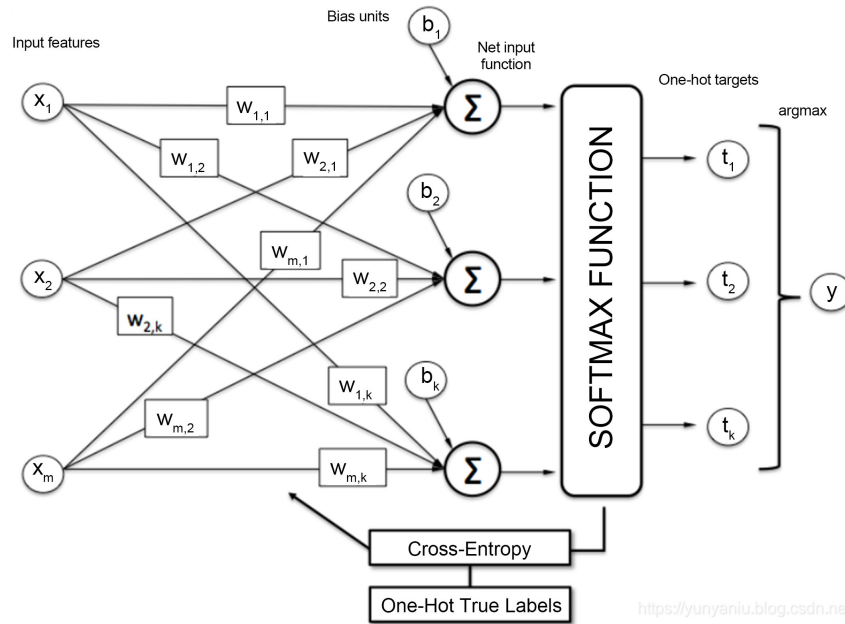


Figure 1. Softmax logistic regression model
图 1. Softmax 建模流程

Table 2. Regression model accuracy results
表 2. 回归模型精度结果

放牧强度	精确率	召回率	调和均值
0	0.51	0.47	0.49
1	0.55	0.38	0.51
2	0.48	0.48	0.48
3	0.65	0.50	0.56

2.2. LSTM 模型

针对解决统计数据预测草原监测样地(12个放牧小区)在不同放牧强度下2023年土壤同期有机碳、无机碳、全碳、全N、土壤C/N比等值的问题。首先建立不同放牧强度和土壤化学性质数学模型，进而通过模型去预测2023年土壤化学性质数据。

LSTM模型，本质上是一种特定形式的循环神经网络(RNN)，主要是为了解决长序列训练过程中的梯度爆炸或梯度弥散问题[3][4]。简单来说，就是相比普通的RNN，LSTM能够在更长的序列中有更好的表现。LSTM模型在RNN模型的基础上通过增加门限来解决RNN短期记忆的问题，使得循环神经网络能够真正有效地利用长距离的时序信息。

2.2.1. LSTM 建模

LSTM在RNN的基础结构上增加了输入门限(Input Gate)、输出门限(Output Gate)、遗忘门限(Forget Gate)3个逻辑控制单元，且各自连接到了一个乘法元件上，其概念图如下，通过设定神经网络的记忆单

元与其他部分连接的边缘处的权值控制信息流的输入、输出以及细胞单元(Memory cell)的状态。其具体结构如下图 2 所示。

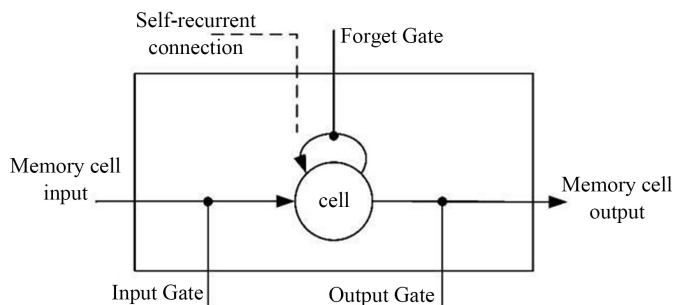


Figure 2. LSTM conceptual diagram

图 2. LSTM 概念图

上图中相关部件的描述如下：

Input Gate: 用于控制信息是否流入记忆单元

Forget Gate: 用于控制上一时刻记忆单元中的信息是否积累到当前时刻的记忆单元中

Output Gate: 用于控制当前时刻的输出值，通过使用激活函数来评估当前输入和记忆状态的组合

cell: 记忆单元，即神经元状态的记忆，它使得 LSTM 单元具有保存、读取、重置和更新长距离信息的能力。

隐藏层 cell 结构如下图 3 所示，在 LSTM 神经网络的训练过程中，首先将 t 时刻的数据特征输入至输入层，经过激励函数输出结果。将输出结果、 $t-1$ 时刻的隐藏层输出和 $t-1$ 时刻 cell 单元存储的信息输入 LSTM 结构的节点中，通过 Input Gate, Output Gate, Forget Gate 和 cell 单元的处理，输出数据到下一隐藏层或输出层，输出 LSTM 结构节点的结果到输出层神经元，计算反向传播误差，更新各个权值。

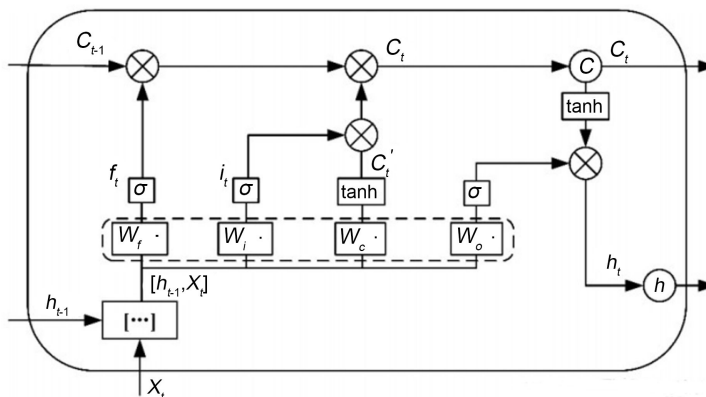


Figure 3. LSTM detailed diagram

图 3. LSTM 细节图

2.2.2. LSTM 预测结果

进行预测锡林郭勒草原监测样地(12 个放牧小区)在不同放牧强度下 2023 年土壤同期有机碳、无机碳、全碳、全 N，而土壤 C/N 比值可以通过预测所得的数据进行解算可得。LSTM 模型原理是：通过多年份同放牧强度和放牧小区的土壤化学性质数据去预测未来某年的同条件下的土壤化学性质数据，并不断回头改进模型准确性。

以 2014~2022 年放牧小区为：G20、放牧强度为：重度放牧强度为例，展示其 LSTM 模型训练的准确度。模型的训练集和测试集比例为 7:3。其训练结果如图 4 所示，其土壤化学性质均方根误差均在 0.31 以下，模型准确(表 3)。

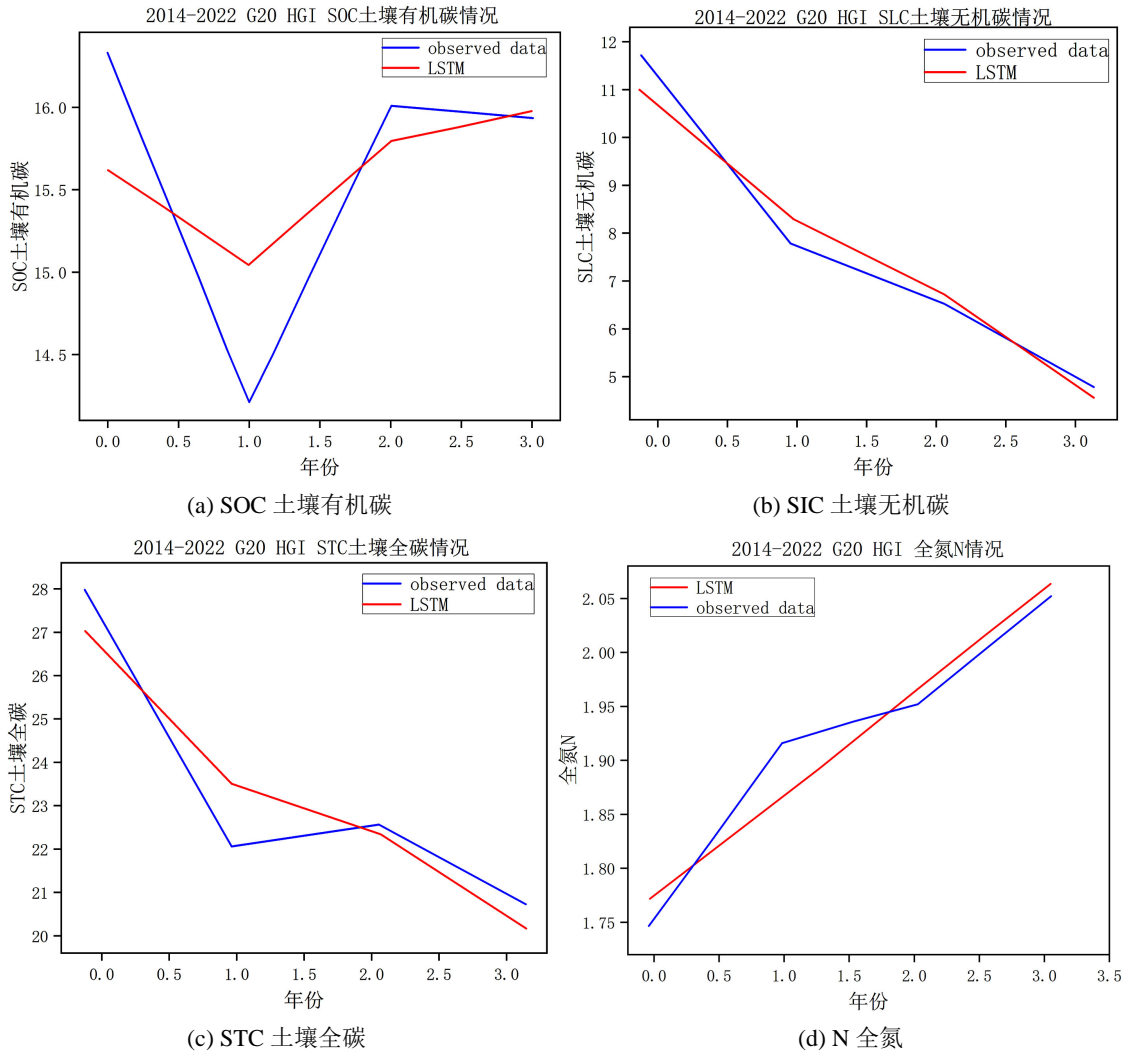


Figure 4. LSTM training results

图 4. LSTM 训练结果

Table 3. Training error of soil chemical properties under heavy grazing intensity in the G20 region

表 3. G20 区重度放牧强度的土壤化学性质训练误差

	有机碳	无机碳	全碳	全氮	碳氮比
RMSE	0.02	0.31	0.28	0.02	0.07
MAE	0.13	0.29	0.25	0.01	0.06
MAPE	1.23	4.16	1.12	0.75	0.52

采用以上 LMST 所得准确模型对 2023 年不同放牧强度和小区的土壤同期有机碳、无机碳、全碳、全 N 进行预测，计算土壤 C/N 比并完成下表 4。

Table 4. Soil chemical properties prediction results
表 4. 土壤化学性质预测结果

放牧强度	Plot	SOC	SIC	STC	全 N	C/N 比
NG	G17	16.85	6.44	23.09	2.12	10.93
	G19	15.50	6.011	21.32	1.94	11.05
	G21	19.47	4.23	23.85	2.34	10.15
LGI	G6	14.58	3.97	18.81	1.90	9.76
	G12	15.05	4.34	20.04	1.86	10.42
	G18	19.06	8.02	26.77	2.24	12.05
MGI	G8	14.38	2.99	17.14	1.94	8.97
	G11	14.25	4.39	18.91	1.86	10.04
	G16	15.17	10.85	25.82	1.62	16.02
HGI	G9	17.43	4.66	20.71	2.33	9.48
	G13	18.12	4.61	21.41	2.20	10.31
	G20	15.65	6.19	21.85	1.99	10.98

3. 沙漠化程度指数与土壤板结化定义

首先需要利用沙漠化程度指数预测模型和数据来确定不同放牧强度下监测点的沙漠化程度数值。通过查阅沙漠化预测公式的来源，确定各指标因子，并查阅锡林郭勒盟年鉴获取缺失的因子数据，计算因子强度，使用层次分析法计算各因子权重，从而计算不同放牧强度下的沙漠化程度指数。

使用因子分析法，对可能的造成土壤板结化的成分进行综合归纳，对各成分提出相应解释从而量化土壤的板结化指数，根据量化模型寻找使土壤板结化程度最小的放牧策略。

3.1. 沙漠化程度指数预测模型

3.1.1. 数据补充与计算因子强度 Q_i

沙漠化的相关影响因素如图 5 所示，有气象因素：风速、降水、气温；地表因素：植被盖度、地表水资源、地下水位；人文因素：人口数量、牲畜数量、社会经济水平。将平均气温作为气温数据、将平均风速作为风速数据、将降水量作为降水数据；由于年鉴中地表水资源和地下水位的每年资料不全，这里将 200 cm 湿度作为地下水位数据、将 10 cm 湿度作为地表水资源数据、将植被指数作为植被盖度数据；通过查阅年鉴数据，将每年的总人口数量作为人口数量数据、将每年 6 月末及 12 月末的牲畜数量作为牲畜数量数据、将每年的全体居民人均可支配收入作为社会经济水平数据。

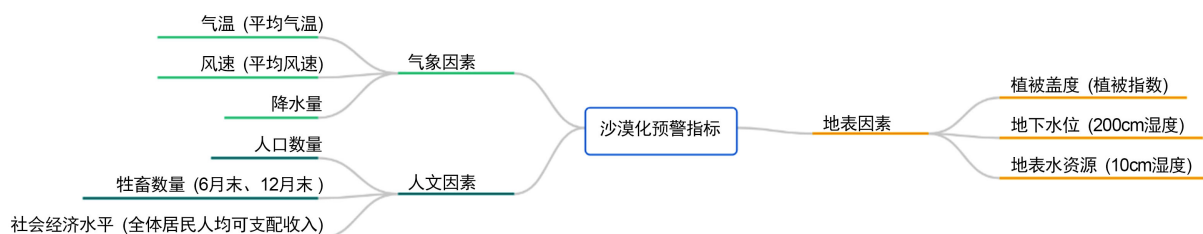


Figure 5. Factors influencing desertification early warning indicators

图 5. 沙漠化预警指标影响因素

通过表 5 因子强度上限和下限以及因子强度(Q_i)的计算方法,对附件中的数据进行计算,将其投射为 0~1 的归一化数据,表 6 展示了部分因子强度 Q_i 的计算结果。

Table 5. Upper and lower limits of factor intensity

表 5. 因子强度的上限和下限

因子	上限	下限	因子	上限	下限
平均风速(m/s)	2.61	1.43	地下水位(m)	7.7	4.85
降水量(mm)	100	10	人口密度(person/km ²)	60	7
平均气温(°C)	35	5	牲畜密度(sheep unit/km ²)	25	5
植被盖度	50	10	人均家庭经营纯收入(yuan)	1100	450

Table 6. Partial calculation results of factor intensity Q_i

表 6. 部分因子强度 Q_i 计算结果

10 cm 湿 (kg/m ²)	200 cm 湿度 (kg/m ²)	植被 指数 (NDVI)	平均气 温(°C)	降水 量 (mm)	平均风 速(m/s)	人口密 度 (person/km ²)	牲畜密 度 (sheep unit/km ²)	社会经 济 水平(元)
0.0731	0.0714	0.3239	0.0532	0.0071	0.3500	0.0000	0.1990	0.0000
0.0638	0.0662	0.3286	0.0564	0.0040	0.3530	0.0032	0.2643	0.0181
0.0601	0.0606	0.3372	0.0599	0.0043	0.2831	0.0086	0.3657	0.0368
0.0640	0.0567	0.3164	0.0582	0.0135	0.2697	0.0151	0.3709	0.0545
0.0680	0.0533	0.2703	0.0572	0.0121	0.2914	0.0231	0.4238	0.0726

3.1.2. 层次分析法计算因子权重 W_{ci}

层次分析法(AHP)将需要决策的复杂问题分解成不同层次和若干因素,通过两两比较确定层次中诸因素的相对重要性,是一种定性和定量相结合的多目标决策方法[5] [6]。土壤板结化属于多目标、多准则以及无结构特性的复杂决策问题,因此,可以应用该方法解决。通过 AHP 构建判断矩阵结果如表 7,最终计算相关因素因子权重如下表 8~10 所示。

Table 7. Three-Factor decision matrix

表 7. 三因素判断矩阵

	气象因素	地表因素	人文因素	特征向量	权重值(%)
气象因素	1	0.3333	5	1.1856	27.8955
地表因素	3	1	7	2.7589	64.9118
人文因素	0.2	0.1429	1	0.3057	7.1927

Table 8. Weights of meteorological factor factors

表 8. 气象因素因子权重

	气温(平均气温)	风速(平均风速)	降水量	特征向量	权重值(%)
气温(平均气温)	1	0.2	0.5	0.4642	10.2825
风速(平均风速)	5	1	8	3.42	75.762
降水量	2	0.125	1	0.63	13.9555

Table 9. Weights of surface factor factors
表 9. 地表因素因子权重

	植被盖度 (植被指数)	地下水位 (200 cm 湿度)	地表水资源 (10 cm 湿度)	特征 向量	权重值(%)
植被盖度 (植被指数)	1	6	8	3.6342	76.1158
地下水位 (200 cm 湿度)	0.1667	1	0.25	0.3467	7.2609
地表水资源 (10 cm 湿度)	0.125	4	1	0.7937	16.6233

Table 10. Weights of human factor factors
表 10. 人文因素因子权重

	人口 数量	牲畜数量 (6 月末、12 月末)	社会经济水平 (全体居民人均可支配收入)	特征 向量	权重值 (%)
人口数量	1	0.1429	0.25	0.3293	7.0383
牲畜数量(6 月末、12 月末)	7	1	6	3.4767	4.2912
社会经济水平 (全体居民人均可支配收入)	4	0.1667	1	0.87361	8.6705

3.1.3. 沙漠化程度指数计算结果

沙漠化程度指数预测模型表达式：

$$SM = \eta \cdot \sum_{i=1}^n S_{Q_i} = \eta \cdot \sum_{i=1}^n (Q_i \cdot W_{ci}) \quad (3)$$

式中： SM 表示沙漠化程度指数； η 为调节系数，用来修正模型； n 为模型中因子的个数； Q_i 为第 i 个因子的因子强度； S_{Q_i} 为第 i 个因子对沙漠化程度的贡献值； x_1, x_2, x_3 为因子权重系数，即因子对沙漠化程度的贡献值等于因子强度与因子权重系数的乘积。

根据 AHP 法得到的各因素因子权重去计算沙漠化程度指数，下表 11 仅展示部分数据。

Table 11. Desertification severity index
表 11. 沙漠化程度指数

年份	月份	沙漠化程度指数
2023	1	0.0011
	2	0.0699
	3	0.3789
2022	1	0.0678
	2	0.1257
...
2014	8	0.9551
	9	0.7354
	10	0.4789
	11	0.3797
	12	0.0801

3.2. 土壤板结化定量分析

土壤板结化与土壤有机物、土壤湿度和土壤的容重有关，目前还没有明确的定量表达式，其数学模型可定性描述为如下：

$$B = f(W, C, O) \quad (4)$$

土壤湿度 W 越少，容重 C 越大，有机物含量 O 越低，土壤板结化程度 B 越严重。

性质和物理性质是影响土壤肥力重要因素，土壤化学性质包括：土壤有机碳 SOC、土壤无机碳 SIC、土壤全碳 STC、全 N、土壤 C/N 比等；土壤物理性质包括：土壤湿度、土壤容重等。

使用因子分析将这些错综复杂的变量对影响土壤板结化关系归结成几个综合因子，并确定其模型。

3.2.1. 因子分析

假设大小为 $n \times p$ 的随机向量 $x = (x_1, x_2, \dots, x_p)'$ 的均值 $u = (u_1, u_2, \dots, u_p)'$ ，协方差矩阵 $\Sigma_{p \times p} = (\sigma_{ij})$ ，因子分析的一般模型为：

$$\begin{cases} x_1 = u_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_2 = u_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \vdots \\ \vdots \\ \vdots \\ x_p = u_{p1}f_1 + a_{p2}f_2 + \dots + a_{pm}f_m + \varepsilon_p \end{cases} \quad (5)$$

其中 $f = (f_1, f_2, \dots, f_m)'$ ($m \leq p$) 为公共因子， $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)'$ 为特殊因子，两者皆为无法观测的随机变量。同时根据 $x = u + Af + \varepsilon$ 构造如下式假设：

$$\begin{cases} E(f) = 0 \\ E(\varepsilon) = 0 \\ \text{Var}(f) = I \\ \text{Var}(\varepsilon) = D = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_e^2) \\ \text{cov}(f, \varepsilon) = E(f\varepsilon') = 0 \end{cases} \quad (6)$$

公因子其具有单位方差且彼此不相关，特殊因子则不仅自身彼此不相关，与公因子也不相关。

3.2.2. 因子分析结果

结合图 6 碎石图和表 12 成分总方差数据可知，影响土壤板结化的前四个因子影响占比高达 85.163%，其影响数值足以用四个成分代表整体，故这里选用四个未知成分去解释土壤板结化模型。

确定的成分数量已确定，然后针对土壤有机物、土壤湿度、土壤的容重以及沙漠化程度指数相关因子进一步获得其影响土壤板结化成分得分系数矩阵，如表 13 所示，每种成分因子所占四个主成分权重可以知道其归属于第几主成分，第一主成分(SOC 土壤有机碳、全氮 N、土壤 C/N 比)；第二主成分(10 cm 土壤湿度、40 cm 土壤湿度、100 cm 土壤湿度)；第三主成分(SIC 土壤无机碳、STC 土壤全碳)；第四主成分(200 cm 土壤湿度、沙漠化程度指数)。

进一步可以对四个成分进行解释，分别为有机量、含水量、含碳量、干燥度。

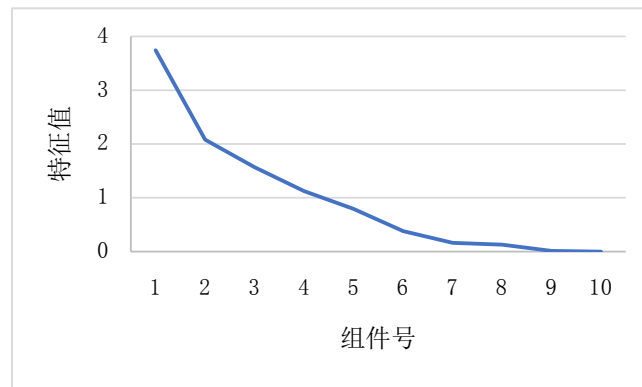


Figure 6. Gravel map

图 6. 碎石图

Table 12. Total variance of components data

表 12. 成分总方差数据

成分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%	总计	方差百分比	累积%
1	3.742	37.418	37.418	3.742	37.418	37.418	3.358	33.576	33.576
2	2.081	20.805	58.223	2.081	20.805	58.223	2.131	21.307	54.883
3	1.569	15.694	73.917	1.569	15.694	73.917	1.556	15.560	70.443
4	1.125	11.246	85.163	1.125	11.246	85.163	1.472	14.720	85.163
5	1.794	7.943	93.106	0.533	6.023	97.635	1.089	14.741	99.694

Table 13. Component score coefficient matrix

表 13. 成分得分系数矩阵

	成分1	成分2	成分3	成分4
SOC土壤有机碳	0.308	-0.016	0.232	0.004
SIC土壤无机碳	-0.202	0.035	0.338	0.047
STC土壤全碳	0.081	0.026	0.675	0.063
全氮N	0.320	-0.007	0.154	-0.059
土壤C/N比	-0.270	0.017	0.160	0.086
10 cm湿度(kg/m ²)	0.000	0.361	-0.005	-0.177
40 cm湿度(kg/m ²)	-0.037	0.427	0.035	0.080
100 cm湿度(kg/m ²)	-0.018	0.380	0.036	0.020
200 cm湿度(kg/m ²)	0.011	-0.100	0.030	-0.471
沙漠化程度指数	-0.098	-0.104	0.125	0.703

3.2.3. 定量模型

通过上述因子分析的结果，可以通过前四主成分占比总量去调整其权重：第一主成分(0.4741)；第二主成分(0.258)；第三主成分(0.1569)；第四主成分(0.111)。四个成分已做说明，分别为有机量、含水量、

含碳量、干燥度。故土壤板结化定量模型如下式：

$$B = 0.4741O_c + 0.258M_w + 0.1569M_c + 0.111D_w \tag{7}$$

式(7-5)中： B 表示土壤板结化； O_c 表示有机量； M_w 表示含水量； M_c 表示含碳量； D_w 表示干燥度。

3.3. 放牧策略模型结果

通过上述得到的沙漠化程度指数模型和量化的土壤板结化指数模型进行联立，见下式：

$$\begin{cases} SM = \eta \cdot \sum_{i=1}^n S_{Q_i} = \eta \cdot \sum_{i=1}^n (Q_i \cdot W_{ci}) \\ B = 0.4741O_c + 0.258M_w + 0.1569M_c + 0.111D_w \\ \{C_l\} = \min \{SM + B\} \end{cases} \tag{8}$$

式中： $\{C_l\}$ 表示使沙漠化程度指数与板结化程度最小的相关因素的集合。通过模型求解最终可获得 $\{C_l\}$ 相关集合，提取集合中的放牧小区和放牧强度因素，最终确定使土壤板结化指数最小的放牧策略见下表 14：

Table 14. Grazing strategy results

表 14. 放牧策略结果

放牧策略	
放牧小区	G16
放牧强度	MGI

3.4. LSTM 预测

要预测放牧区 2023 年九月的土地状态，衡量土地状态的指标有土壤肥力变化、土壤湿度、植被覆盖等。此处使用对土壤板结化进行因子分析后综合得出的指标来替代，将有机量作为土壤的肥力，将含水量作为土壤湿度，将叶面积指数作为植被覆盖。沿用上面的 LSTM 建模方法，综合 2014~2022 年九月的土壤数据，使用 LSTM 法进行学习并 2023 年九月的土地状态进行预测。

预测结果

2014~2022 年九月有机量、含水量和叶面积指数 LSTM 模型的训练集和测试集比例为 7:3，其训练结果如表 15、图 7 所示。三种土地状态指标的 RMSE 值均在 0.62 以下、MAE 值不超过 4.5，因此模型准确。

Table 15. Model training error

表 15. 模型训练误差

土地状态指标	RMSE	MAE
有机量	0.56	0.28
含水量	0.62	0.45
叶面积指数	0.25	0.15

通过上述所建立的 LMST 模型进行 2024 年九月的土地状态的预测，如图 8 所示。蓝色点表示 2014~2022 年九月的土地状态，绿色点表示这几年的平均值，红色点为预测得到的 2024 年九月的土地状态。可以看出 2024 年的土壤，其植被盖度和土壤湿度得到了改善，而相应的，土壤的肥力会下降。

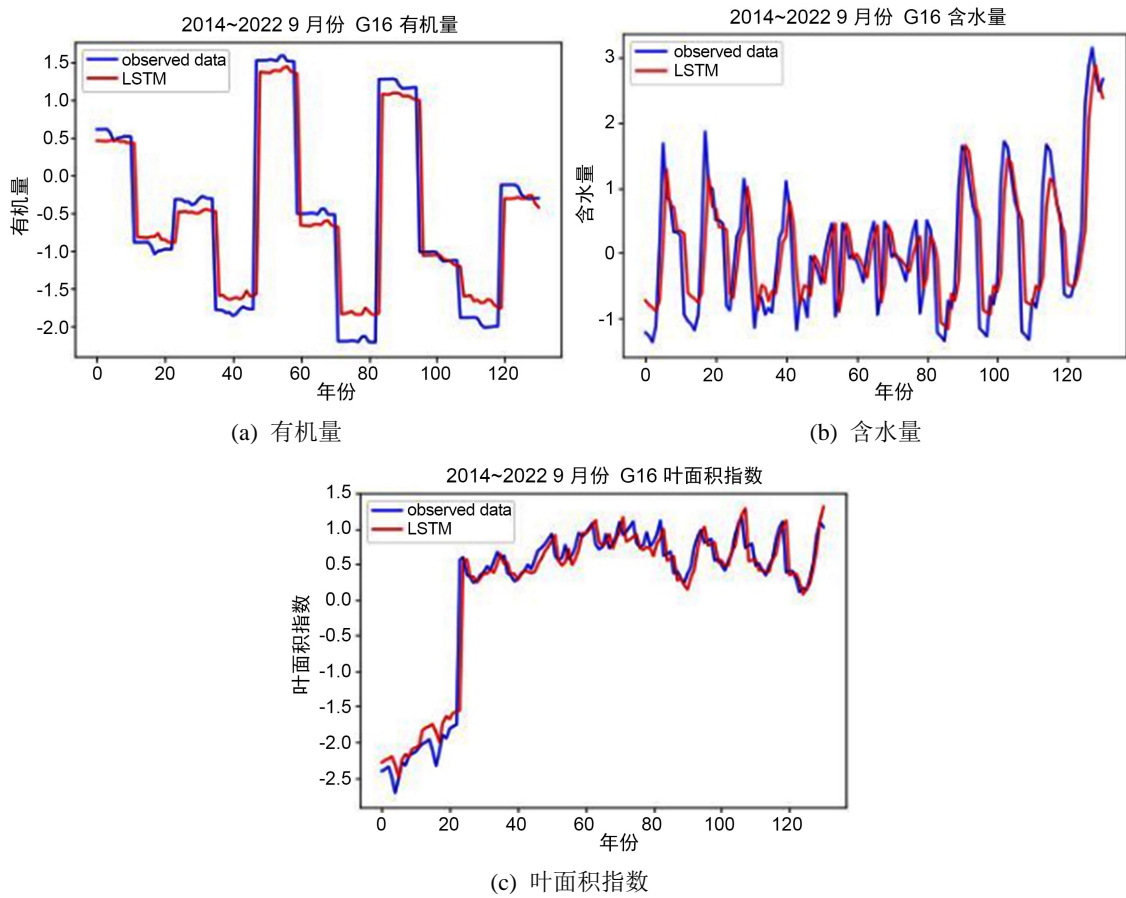


Figure 7. LSTM model training results
 图 7. LSTM 模型训练结果

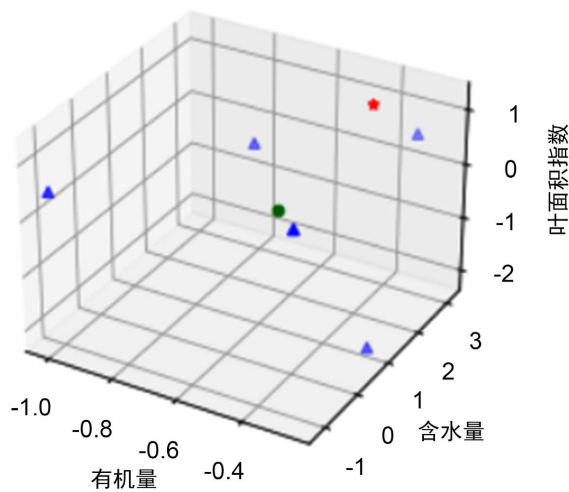


Figure 8. Soil condition in September 2024
 图 8. 2024 年 9 月的土壤状态

4. 结论

综合以上分析，本研究通过 Softmax 逻辑回归模型和 LSTM 模型，成功建立了草原土壤状态与放牧

策略的预测模型。通过对历年统计年鉴数据的深入分析,我们得出了六种土壤状态在不同放牧策略下的四分类模型,并通过 LSTM 模型预测了 2023 年的土壤状态数据。沙漠化程度指数的运用进一步确定了监测点在不同放牧强度下的沙漠化程度数值。最后,我们采用了有机量、含水量、叶面积指数等多指标代替传统的土壤肥力、土壤湿度、植被覆盖等指标,综合分析 2014~2022 年的土壤数据,并通过 LSTM 模型成功预测了 2024 年同月的土壤状态数据。这些预测结果为草原的可持续保护和科学合理开发提供了重要的参考,为未来的草原管理和决策制定提供了可靠的数据支持。

参考文献

- [1] 邱倩倩, 张卓栋, 孙传龙, 刘宏远. 锡林郭勒草地景观系统土壤容重空间变异及其与风蚀的关系[J]. 水土保持通报, 2016, 36(6): 58-62+66.
- [2] 邓国伦, 邓振镛, 万信, 方德彪. 陇东黄土高原土壤贮水量对小麦产量影响的研究[J]. 气象科技, 1998, 26(1): 56-60.
- [3] 郭新伟, 王彩玲. 基于 LSTM 模型的黄河宁蒙河段水质预测研究[C]//2023 (第十一届)中国水利信息化技术论坛论文集. 包头: 黄河水利委员会宁蒙水文水资源局, 2023: 10.
- [4] 李丹青, 杨超宇. 基于 ARIMA 和 LSTM 的瓦斯浓度预测研究[J]. 新乡学院学报, 2022, 39(9): 45-48.
- [5] 方雅青. 医务社会工作者绩效评价指标体系研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2012.
- [6] 刘妍. A 高校校园超市供应商评价体系研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2020.