

袁世一. 基于数据分解与改进混合深度网络的大豆产量预测[J]. 智能计算机与应用, 2025, 15(6): 27-31. DOI: 10.20169/j.issn.2095-2163.250605

基于数据分解与改进混合深度网络的大豆产量预测

袁世一

(中国农业科学院 农业信息研究所, 北京 100081)

摘要: 大豆因产量数据较少且受外部因素影响较大,其产量预测存在一定困难。为解决这一问题,本文从数据处理和模型优化角度出发,建立了基于数据分解和改进的混合深度网络的大豆产量预测模型。首先,利用改进后的自适应集成经验模态分解算法对大豆数据进行模态分解,降低数据波动性。其次,建立混合神经网络预测模型,提高预测的准确性和稳定性。最后,根据 APSO 算法提取各模型及各分量中的优势,将提取后的数据进行重构,进一步提高模型的预测精度。结果表明,本文所提出的算法改进模型,在大豆产量预测上具有更高的精度。

关键词: 数据分解; 混合深度网络; 大豆产量; 预测

中图分类号: TP181

文献标志码: A

文章编号: 2095-2163(2025)06-0027-05

Soybean yield prediction based on data decomposition and improved mixed depth network

YUAN Shiyi

(Institute of Agricultural Information, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: It is difficult to forecast soybean yield because of the small amount of data and the great influence of external factors. In order to solve this problem, a soybean yield prediction model based on data decomposition and improved mixed depth network is established from the perspective of data processing and model optimization. Firstly, the improved adaptive integrated empirical mode decomposition algorithm is used to decompose soybean data to reduce data fluctuations. Secondly, a hybrid neural network prediction model is established to improve the accuracy and stability of the prediction. Finally, according to the APSO algorithm, the advantages of each model and each component are extracted, and the extracted data is reconstructed to further improve the prediction accuracy of the model. The results show that the improved model proposed in this paper has higher accuracy in soybean yield prediction.

Key words: data decomposition; hybrid deep network; soybean yield; forecast

0 引言

当前,国内大豆产量供需矛盾依然存在。因此对大豆产量进行精准预测,不仅能为科学调控提供有力的数据支撑,同时也能以此为基础来激发农业经济生产的积极性,从而保障国内粮食安全稳定性。因此,如何实现大豆产量的精准预测至关重要^[1]。

针对大豆产量预测主要可分为常规统计学模型^[2]与机器学习^[3]。在统计学模型方面,学者们主要通过专家经验或简单的数学模型,对大豆的历史数据进行拟合。部分学者从大豆影响因素,如气象因素、历史产量等方面进行考虑,建立了大豆产量预

测模型。但是大豆产量的数据相对复杂,其数据具有较强的波动性,传统统计学方法难以精准地实现大豆产量的预测,存在顶峰误差较大的情况。随着机器学习算法的优势不断凸显,由此实现的数据挖掘与数据处理方法正陆续应用于大豆产量预测的研究中。在大豆产量预测上,机器学习方法、如支持向量机和 BP 神经网络具有一定的效果,但也存在一些不足。其中,支持向量机训练时间长,模型核函数的选择和最佳参数的确定对预测结果有重大影响。而 BP 神经网络在训练过程中可能出现“早熟”现象。为了解决这些问题,本文提出了一种新的研究

基金项目: 国家自然科学基金(62103418); 中国农业科学院农业信息研究所基本科研业务(JBYW-AII-2024-30)。

作者简介: 袁世一(1988—),女,博士,助理研究员,主要研究方向:粮食安全监测预警。Email:yuanshiyi@caas.cn。

收稿日期: 2023-11-13

哈尔滨工业大学主办 ◆ 学术研究与应用

是为了解决 RNN 存在的梯度爆炸和梯度消失问题。这 2 种网络的核心是在信息流动过程中添加了控制机制,以保护和控制信息的传递和更新。GRU 网络的内部结构主要包括 2 个门、即重置门和更新门,这使得其结构相较于 LSTM 更为简洁,并且计算效率更高。重置门负责控制过去的信息流向,而更新门则用于控制新的输入信息和过去信息的融合程度。本文运用的 GRU 网络原理及结构参照相关文献 [8]。

1.4 适应性粒子群算法

粒子群算法(PSO)因其简明规则和对优化问题的强大求解能力而被广泛利用。但是,由于该算法通常存在过早收敛和多样性丧失的问题,这可能阻碍获得精确的最优结果^[9-11]。适应性粒子群算法(APSO)通过适应度选择加速系数的方式,成功消除了 PSO 算法中随机加速系数的影响,从而克服了 PSO 的局限性,使得获取精确解成为可能。然而,APSO 依然面临粒子过早收敛的问题,即在优化求解过程中,粒子跳出局部最优解的能力仍然有限。此外,APSO 在初始化阶段使用随机分布序列来产生粒子的初始位置,这可能降低粒子种群的均匀性和多样性。为了解决这些问题,本文提出了一种改进的粒子群算法。首先,使用 Piecewise 映射来初始化粒子种群,以提高种群的均匀性和多样性^[12]。其次,采用 Levy 飞行策略进行粒子位置的二次更新,以增强算法寻找全局最优解的能力。改进后的粒子群算法的计算流程如图 2 所示。

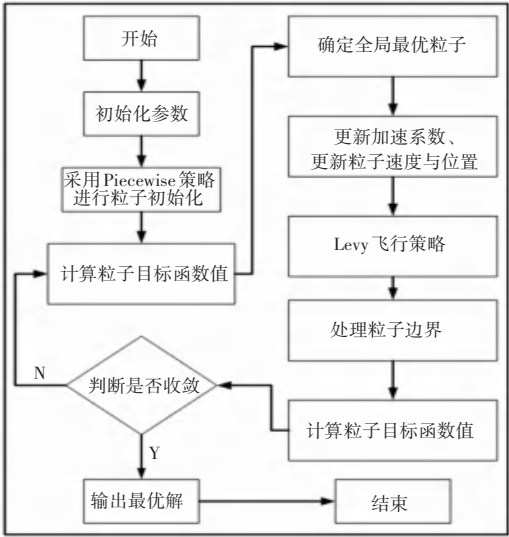


图 2 改进后的粒子群算法求解流程

Fig. 2 Solution flowchart of the improved particle swarm optimization algorithm

2 实验评估

2.1 仿真基本数据

为了验证本文所提模型的合理性与预测结果的准确性,以 1949~2022 年中国大豆年产量为基本数据,建立大豆产量预测模型。其中,将前 80%数据作为预测集,后 20%数据作为测试集。1949~2022 年,国内大豆年产量整体呈上升趋势。依据上述基础数据及模型训练、预测集划分方式,建立基于数据分解与改进混合深度网络的大豆产量预测模型。

2.2 基于改进的自适应集成经验模态分解算法结果分析

根据大豆产量基本数据可知,训练集中大豆年产量数据存在一定的波动性,同时预测模型整体数据量较少。为了提升预测模型的特征学习及数据挖掘能力,降低数据波动性对网络预测结果的影响,本文利用改进后的自适应集成经验模态分解算法对大豆年产量数据进行模态分解,将其分解为多个相对平滑的本征模态分量。

根据分解结果可知,在生成第一个本征模态分量(IMF1)时,分解后数据波动较为严重,而后续本征模态分量的波动情况明显更趋平缓,其中 IMF4 的波动情况最小,曲线更加平滑。将上述分解后得到的各时间点下大豆产量本征模态预测结果作为深度组合网络的基本数据,建立基于 APSO 算法优化的组合深度神经网络分量预测模型,并将各分量网络预测结果进行重构获取最终大豆产量预测结果。

2.3 基于改进混合深度网络大豆产量预测结果分析

根据大豆产量分解结果,本文利用基于 APSO 算法优化的混合深度网络实现大豆产量的精准预测。其中,将前 80%数据作为训练集,后 20%数据作为测试集。采用该混合深度网络进行预测后,各分量神经网络训练集预测结果如图 3 所示。

在训练集预测结果中,各 IMF 分量预测结果处 IMF1 分量外其余分量预测趋势均较为准确。其中,IMF4 预测结果准确性更高,而 IMF1 分量预测模型的预测精度最低。这与大豆产量模态分解结果中各模态分量的波动情况相对应,由此可见数据波动性对模型预测精度影响较大。

混合神经网络中的 3 种主要神经网络、即 GRU、LSTM 及 BP 神经网络在不同 IMF 分量时所表现出的预测能力有一定差异。其中,LSTM 及 GRU 模型在 IMF2 及 IMF3 后段预测中表现得更为准确,尤其是 GRU 预测模型,其整体预测趋势与实际 IMF

分量更加相符。但 LSTM 及 GRU 预测模型在 $IMF4$ 分量中的预测精度均低于 BP 神经网络的预测精度。根据模态分解结果可知, $IMF4$ 的分解结果较为平滑, 数据波动性极小, 因此采用 BP 等浅层神经

网络即可实现较好的预测结果。而 GRU 及 LSTM 等深度学习网络, 受限于模型输入数据较少等因素, 其在初始时刻预测结果较差, 即存在一定的初段误差较大的情况。

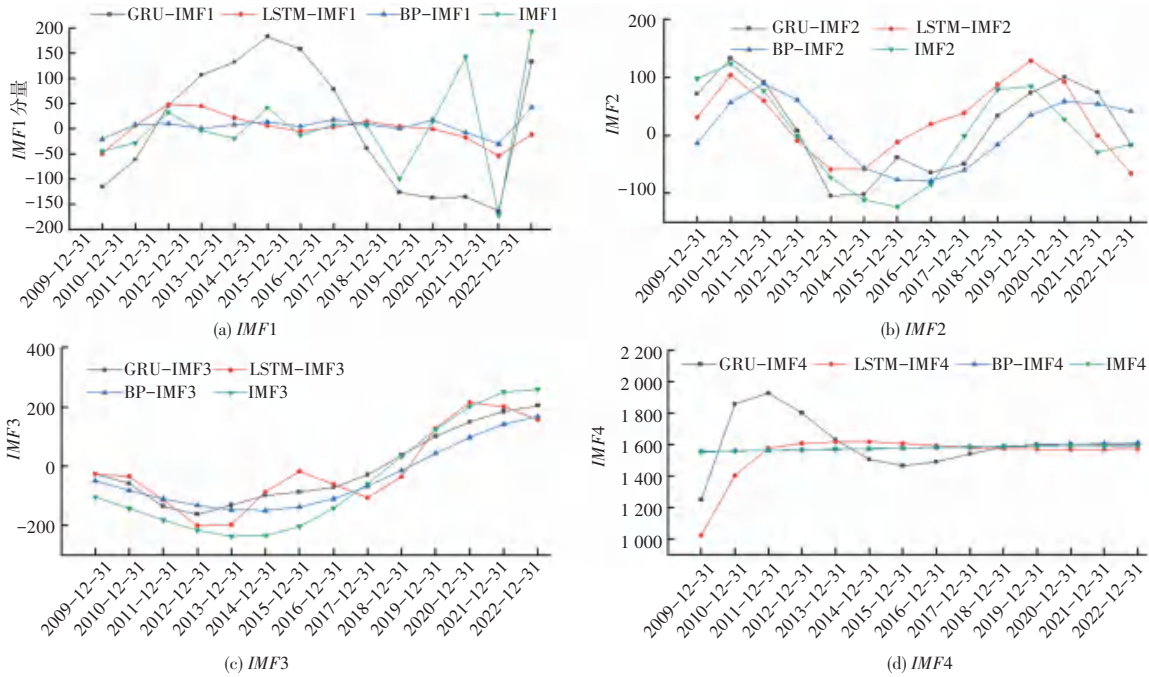


图 3 各分量 IMF 测试集预测结果

Fig. 3 Forecast results of IMF test set for each component

为了避免上述问题的发生, 本文利用 APSO 算法, 结合分量线性权重的方式对各预测模型进行加权寻优, 提取各模型中精度较高的分量作为大豆产量预测数据的主体, 削弱预测模型中较差的部分对预测结果的影响。图 4 为采用 APSO 算法优化后的改进混合深度网络大豆产量预测结果。

性方面的特征学习及数据挖掘能力。

为了进一步体现本文所提模型的先进性及预测结果的准确性, 分别以均方根误差 ($RMSE$)、平均绝对百分比误差 ($MAPE$) 及回归系数 (R) 作为评价指标, 对本文模型进行评价对比。评价结果见表 1。

表 1 不同算法下大豆产量指标对比

Table 1 Comparison of soybean yield indexes under different algorithms

名称	$RMSE$	$MAPE/\%$	R
BP	120.08	6.90	0.100
LSTM	217.36	10.71	0.210
GRU	254.74	15.23	0.760
本文模型	64.20	2.80	0.931

根据表 1 结果可知, 与传统 BP、LSTM 及 GRU 网络相比, 本文所提出的基于数据分解及 APSO 算法改进的混合深度网络模型, 在大豆产量预测上具有更高的精度, 其 $RMSE$ 、 $MAPE$ 及回归系数 R 的指标均有极大的提升。具体来说, $RMSE$ 指标分别降低约: 55.88、153.16 及 190.54。这极大地提升了大豆产量预测在相关政策支撑方面的重要性, 具有一定的实际意义。

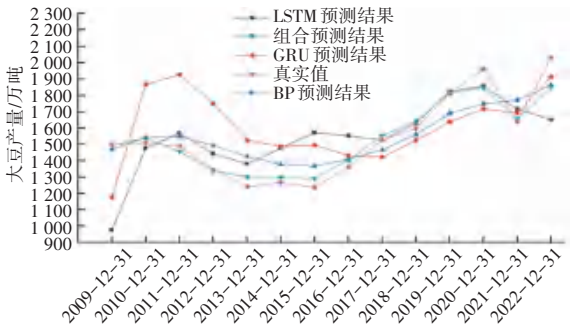


图 4 基于 APSO 优化的混合深度网络大豆产量预测

Fig. 4 Soybean yield prediction based on APSO-optimized mixed depth network

根据图 4 可知, 采用 APSO 算法与混合深度网络相结合后, 大豆产量预测结果有了显著的提升。混合神经网络汲取了 BP 网络及 LSTM、GRU 等深度网络的优点, 克服了深度学习网络对大数据量的依赖, 同时提升了浅层神经网络在数据波动性及复杂

3 结束语

针对大豆产量的精准预测问题,本文提出了一种改进的混合深度网络预测方法。该方法通过采用改进后的自适应集成经验模态分解算法对历史大豆产量数据进行模态分解,建立混合神经网络预测模型,使用 APSO 算法提取各种模型以及分量优势,得到最终的大豆产量预测结果。研究表明:使用 ICEEMDAN 算法能够有效地降低数据波动性对大豆产量预测的影响。同时,对分解后的数据进行分量预测可以提高模型的预测精度。混合神经网络与 APSO 算法相结合的预测方法,能够同时保留深度神经网络和浅层神经网络的优势,极大地提高了大豆产量的预测精度。基于 ICEEMDAN 与 APSO 相结合的改进混合深度网络预测模型,可以实现大豆产量的精确预测。此外,该模型能够降低数据波动性对预测结果的影响。

参考文献

[1] 刘金宇,王兴环,刘洋宸,等. 气象因子对宝清县大豆和玉米产量的影响及预测[J]. 农业灾害研究,2022,12(2):107-109.
[2] 于彦亮,张莉. 2000 年辽宁省粮食及大豆产量的预测:元线性回

归和灰色系统 GM(1,1) 预测模型的应用[J]. 农业系统科学与综合研究,1992(2):129-130.
[3] 张佩,陈郑盟,马顺登,等. 用冠层叶色偏态分布模式 RGB 模型预测大豆产量[J]. 农业工程学报,2021,37(9):120-126.
[4] 陈家栋,田宏伟,黄进,等. 基于农业灾情 EEMD 的河南省干旱时空演变特征[J]. 水土保持研究,2023,30(5):250-256.
[5] 冯亮,李昕,史宏达,等. 波浪能发电平台系泊系统耦合动力响应及水动力分析[J]. 中国海洋大学学报(自然科学版),2019,49(3):171-178.
[6] 韩红,冯光辉,邓国取. 基于 ND-GAIN 模型的河南省农业抗灾能力评价[J]. 水利水电科技进展,2022,42(1):67-74.
[7] 曹英丽,林明童,郭忠辉,等. 基于 Lab 颜色空间的非监督 GMM 水稻无人机图像分割[J]. 农业机械学报,2021,52(1):162-169.
[8] GUO Xifeng, WANG Kaize, YAO Shu, et al. RUL prediction of lithium ion battery based on CEEMDAN-CNN BiLSTM model[J]. Energy Reports, 2023, 9(10): 1299-1306.
[9] 陈晓东,陈祥,冯勇,等. 改进 PSO 算法在 AWS 波浪发电场参数辨识中的应用[J]. 现代电子技术,2023,46(16):139-142.
[10] 贾俊平. 统计学[M]. 北京:中国人民大学出版社,2018.
[11] CHEN Zhengmeng, WANG Fuzheng, ZHANG Pei, et al. Skewed distribution of leaf color RGB model and application of skewed parameters in leaf color description model[J]. Plant Methods, 2020, 16: 23.
[12] HUANG Xiaolin, MEHRKANOON S, SUYKENS J A K. Support vector machines with piecewise linear feature mapping[J]. Neurocomputing, 2013, 117: 118-127.