

基于时间序列与横截面数据的吉林省 水稻产量预测对比分析*

陈 威¹, 祁伟彦¹, 袁福香², 李哲敏^{1*}

(1. 中国农业科学院农业信息研究所, 北京 100081; 2. 吉林省气象科学研究所, 长春 130062)

摘要:【目的】对比不同模型预测效果, 分析各模型预测水稻产量的特点、不足及适用条件, 为粮食产量预测问题模型选择提供依据。【方法】从时间序列预测和横截面数据预测两种角度, 利用 ARIMA、LSTM、SVR、MLP 这 4 种模型, 通过吉林省水稻产量、病虫害及其他特征历史数据对吉林省水稻产量进行预测, 并对不同模型的预测结果进行了对比分析。【结果】基于 ARIMA 模型和 LSTM 模型的时间序列预测结果较好, 横截面数据预测中, 原始数据经主成分分析 PCA 降维处理后, 可提高模型预测性能。【结论】对于水稻产量预测, 应根据掌握的影响产量因素的数据以及趋势延续性情况合理选择预测模型, 以达到较理想的预测效果。

关键词: 水稻产量; ARIMA; LSTM; SVR; MLP; 时间序列

DOI: 10.12105/j.issn.1672-0423.20180510

0 引言

水稻是世界三大粮食作物之一, 是我国最主要的粮食作物, 全国 65% 以上的人口以稻米为主食。我国水稻播种面积占全国粮食播种面积的 27% 左右, 产量占全国粮食总产量的 37% 左右^[1]。因此, 水稻的稳产增产以及对水稻产量的准确预测对我国农业政策调整 and 保障我国粮食安全问题具有重要意义。

国内学者在粮食产量预测方面做了大量研究, 有的利用粮食产量的时间序列数据进行预测, 有的利用与粮食生产密切相关的多重参数的横截面数据进行预测。时间序列数据预测常用的预测算法有自回归滑动平均、人工神经网络等。由于时间序列预测只需变量本身的历史数据, 不需要其他参数的数据, 模型构建较为简单, 因此在粮食产量预测中应用十分广泛。然而, 时间序列预测假设粮食产量具有线性变化的规律, 但粮食生产具有受多维度因素影响的复杂性, 使其呈现非线性变化。因此, 单纯采用时间序列的方法进行粮食产量预测, 预测精度不高。横截面数据预测常用的方法有支持向量回归、神

收稿日期: 2018-08-10

第一作者简介: 陈威 (1981—), 男, 满族, 湖北黄梅人, 副研究员、硕士生导师。研究方向: 分布式信息检索、农业信息技术、农产品监测预警。Email: chenwei@caas.cn

* 通信作者简介: 李哲敏 (1970—), 女, 汉族, 福建安溪人, 研究员、博士生导师。研究方向: 农业信息分析与预警、食品安全监测预警。Email: lizhemin@caas.cn

* 基金项目: 农业部 2017 年创新人才项目; 中国农业科学院科技创新工程项目 (CAAS-ASTTP-2017-AII-02)

神经网络等。在利用横截面数据进行粮食产量预测以及支持向量回归构建模型时参数一般依靠经验来确定，而神经网络训练模型对网络初始权重高度敏感，且对训练集样本的依赖性强，因此预测结果有时并不理想。

水稻产量受自然因素、生产技术因素、社会经济因素和随机因素等多因素影响，是线形关系和非线性关系并存的复杂性问题。使用单一模型进行水稻产量预测，当影响产量的重要因素发生变化时可能导致预测结果的不稳定。为了对比不同模型预测效果，分析各模型预测水稻产量的特点、不足及适用条件，为粮食产量预测问题模型选择提供依据，该研究从时间序列预测和横截面数据预测两种角度，选取 4 种模型对水稻产量进行预测。在时间序列预测方面，依据吉林省水稻产量的历史数据，基于自回归滑动平均（Autoregressive Integrated Moving Average Model, ARIMA）模型和长短期记忆人工神经网络（Long-Short Term Memory, LSTM）模型进行吉林省水稻产量预测。ARIMA 模型在粮食产量预测中应用广泛^[2-3]，而 LSTM 模型属于反馈神经网络的范畴，在时间序列的学习和处理上有独特优势，在语音及图像识别、机器翻译等方面已有较多研究^[4-6]，但在粮食产量预测中应用较少^[7]。由于农作物生长产生过程中动物害虫和病原体的侵害可引起显著的产量潜在损失^[8-9]，竞争性作物品种种植面积对水稻的产量也有重要影响，因此选取吉林省重要病虫害发生情况、玉米和大豆种植面积产量等历史数据作为横截面数据预测的原始数据。在横截面数据预测模型选择上，选取支持向量回归模型和多层感知器模型进行预测。支持向量回归模型在解决小样本问题中具有优势，已有学者采用此模型进行农作物产量的预测^[10]，而多层感知器模型作为最简单的神经网络模型，具有良好的容错性和自适性，可有效处理非线性可分离问题，在语音识别、机器翻译等方面已有大量研究^[11-13]，用于粮食产量预测的报道不多。

1 模型简介

1.1 自回归积分滑动平均（ARIMA）模型

ARIMA 模型全称为自回归积分滑动平均模型，是指将非平稳时间序列转化为平稳时间序列，然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。ARIMA 模型的 3 个参数为 p 、 d 、 q ，其中 p 代表预测模型中采用的时序数据本身的滞后数，也叫做自回归项数， d 代表时间序列成为平稳时所做的差分次数， q 代表预测模型中采用的预测误差的滞后数，也叫做移动平均项数^[14]。

1.2 长短期记忆人工神经网络（LSTM）模型

长短期记忆人工神经网络是一种改进的循环神经网络（Recurrent Neural Network, RNN），避免了传统 RNN 隐含层层数过多计算量庞大而产生的梯度消失和梯度爆炸的问题。LSTM 的经典结构如图 1 所示，Cell 是计算节点的核心，通过遗忘阀门（forget gate）、输入阀门（input gate）和输出阀门（output gate）的打开或者关闭来判断模型网络的记忆态（之前网络的状态）在该层输出的结果是否达到阈值，从而加入到当前该层的计算中^[15]。

2018 年 10 月

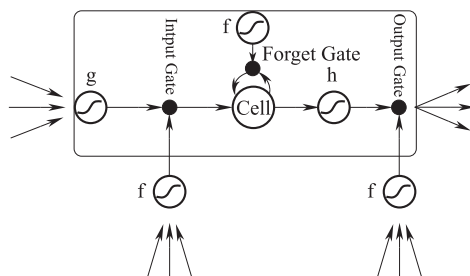


图 1 长短期记忆神经网络 (LSTM) 模型结构

Fig.1 Model structure of LSTM (Long-Short Term Memory)

1.3 支持向量回归 (SVR)

支持向量机 (Support Vector Machine, SVM) 是机器学习领域中基于统计学习理论的分类型回归方法, 由 AT&T 贝尔实验室的 Vapnik 于 20 世纪 90 年代提出, 通过构造核函数将原空间非线性问题转换成高维空间的线性问题, 可以成功处理分离问题, 辨别分析回归问题等^[16]。在处理分类问题时, 支持向量机相当于找到一个高维特征平面, 使两个分类集合的支持向量或者所有数据离分类平面最远, 而支持向量回归 (Support Vector Regression, SVR) 则是找到一个高维回归平面, 使一个集合中的所有数据到该平面的距离最近。

1.4 多层感知器 MLP 模型

MLP (Multilayer Perceptron, MLR), 即多层感知器, 是一种常见的神经网络模型, 由输入层、隐藏层和输出层组成, 其基本结构如图 2 所示。MLP 是一种监督学习模型, 在模型训练过程中, 不断提供完整的输入和输出, 模型基于一定长度的历史数据不断训练、学习, 从而构建出输入和期望的输出之间的最优模型, 而后利用得出的最优模型进行预测^[17]。

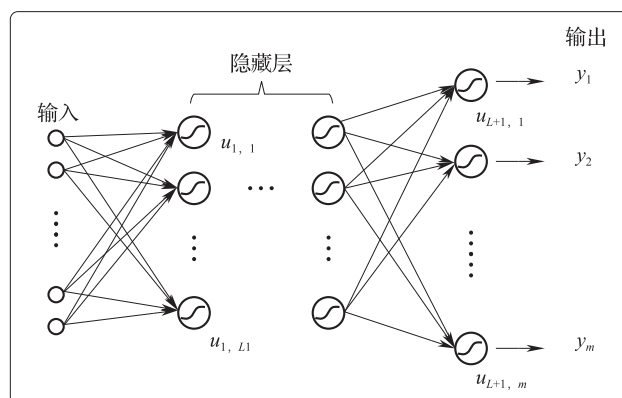


图 2 多层感知器 (MLP) 基本结构

Fig.2 Basic structure of MLP (Multilayer Perceptron)

2 水稻产量预测实证分析

2.1 数据来源

研究中所用到的实验数据为 1949—2015 年吉林省玉米、大豆及水稻的种植面积、单

产及总产量数据，以上数据来自吉林省统计局；吉林省稻瘟病发病频率（1961—2009 年）、玉米螟发生情况（1980—2007 年）、地下害虫及大豆蚜发生情况（1981—2006 年）、食心虫发病情况（1982—2006 年）、黏虫发生情况（1979—2006 年），以上数据来自吉林省气象局。

以 1990 年为基期，1990 年之前的数据作为训练集数据用于模型训练，1990 年之后的数据作为预测集数据，用于检测模型预测效果。

2.2 评价指标

使用均方根误差值（Root Mean Square Error, *RMSE*）作为评价指标对各模型模拟效果进行评价和比较。*RMSE* 是观测值与真值偏差的平方和观测次数 *n* 比值的平方根，计算方法见公式（1）。*RMSE* 值越小，表明预测值与拟合值之间偏差越小，预测性能越好，反之则预测性能越差。

$$RMSE=\sqrt{\frac{\sum_{i=1}^n(X_{obs,i}-X_{model,i})^2}{n}}$$

(1)

2.3 从时间序列角度预测产量

（1）基于 ARIMA 的时间序列预测

①时间序列平稳性检验及平稳化处理

导入吉林省 1949—1990 年历年水稻总产量原始数据，并检验时间序列的稳定性。若时间序列的统计学特性（例如均值、方差）随时间保持不变，则认为该序列是平稳的。



图 3 1950—1990 年吉林省水稻总产量
Fig.3 Total rice yield in Jilin province from 1950 to 1990

如图 3 所示，1949—1990 年吉林省水稻总产量呈上升趋势，年增长幅度不同。由移动平均值和移动标准方差曲线可以看出，移动标准方差随时间变化较小，但移动平均值随时间呈明显的上升趋势，因此该序列不是平稳序列，需要进行平稳化处理。

由于原数据值域范围比较大，为了缩小值域，同时保留其他信息，对原始数据进行对数转换，之后采用差分的方法，将每一年的数值与滞后 10 年的值作差，去除数据的趋势性（图 4）。

2018 年 10 月

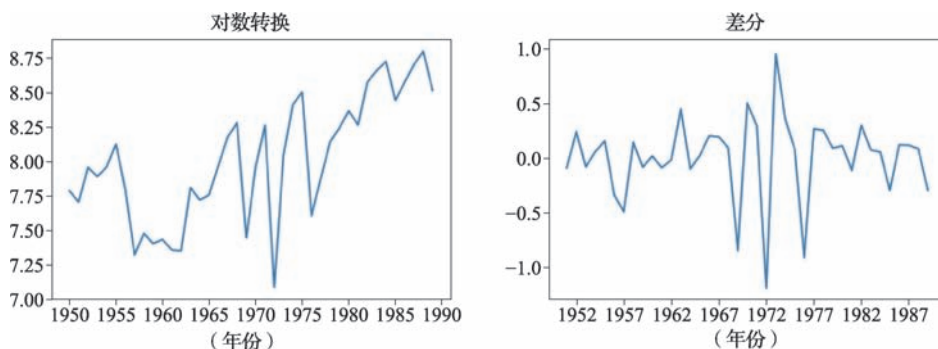


图 4 原始数据平稳化处理

Fig.4 Stationary process of raw data

经对数转换和差分处理后，数据的移动均值和移动标准方差在时间轴上振幅明显缩小，DFtest 结果显示 Test Statistic 的值小于 Critical Value (1%) 的值，表明在 99% 的置信度下，该时间序列数据是稳定的，如图 5。

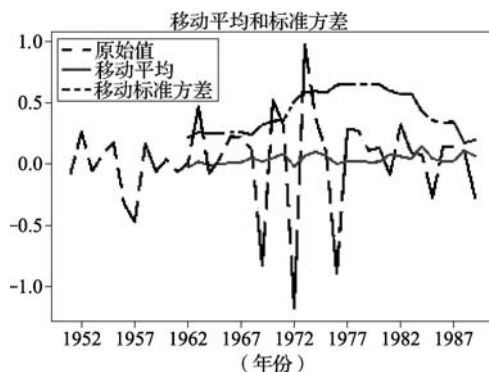


图 5 平稳化处理后移动均值与移动标准方差曲线

Fig.5 Rolling mean and standard deviation after stationary process

②构建 ARIMA 模型

通过自相关函数 (ACF) 和部分自相关函数 (PACF) 进行 ARIMA (p, d, q) 的 p, q 参数估计。由前文差分部分已知，原始数据经一阶差分后数据已经稳定，可得到 $d=1$ 。所以用一阶差分化的 $Y'_t=Y_t-Y_{t-1}$ 作为输入。画出 ACF, PACF 的图像 (图 6)。

图 6 中，上下两条灰线之间为置信区间， p 的值为 ACF 第一次穿过上置信区间的横轴值。 q 的值为 PACF 第一次穿过上置信区间的横轴值。所以从图 6 可以得到 $p=2, q=2$ 。由此，可生成模型 ARIMA (2, 1, 2)。

取 AR 模型 (ARIMA (2, 1, 0))、MA 模型 (ARIMA (0, 1, 2))、ARIMA 模型 (ARIMA (2, 1, 2)) 这 3 种参数取值的 3 个模型与稳定化处理后的输入数据进行拟合，比较 RSS 值得出，ARIMA 模型 (ARIMA (2, 1, 2)) 拟合度最好 (图 7)。

③ ARIMA 模型预测

对 ARIMA 模型 (ARIMA (2, 1, 2)) 拟合值进行差分和对数处理的逆运算 (图 8a)，使拟合值回到原始区间，并利用该模型对 1991—2015 年水稻产量进行预测 (图 8b)。

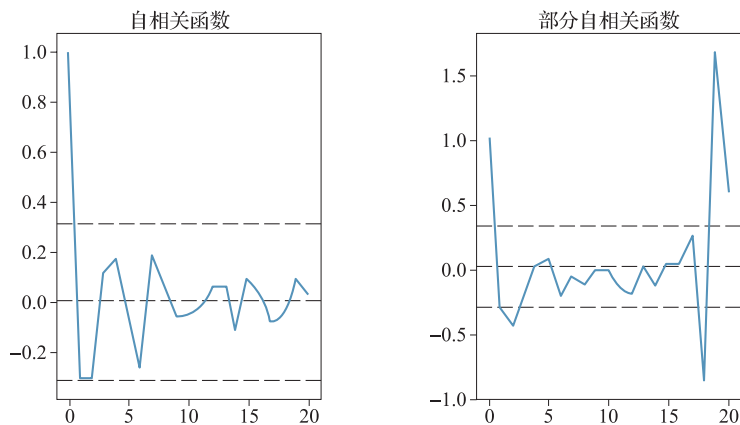


图 6 自相关函数 (ACF) 和部分自相关函数 (PACF) 曲线
Fig.6 Curve graph of ACF and PACF

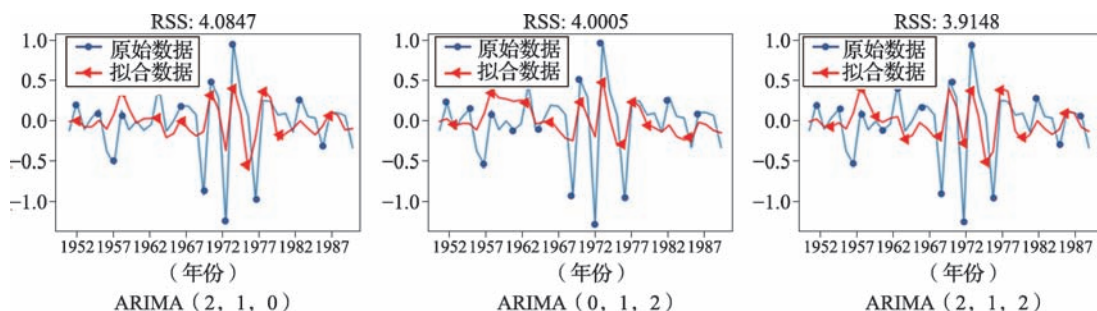


图 7 3 种参数 ARIMA 模型与平稳数据拟合度分析
Fig.7 Fitting analysis of three parameters ARIMA model with stationary data

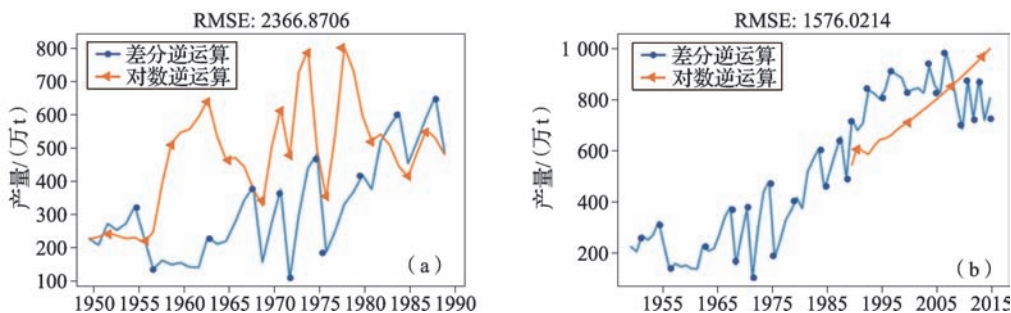


图 8 差分与对数的逆运算 (a) 及水稻产量预测 (b)
Fig.8 Inverse operation of difference and logarithm (a) and yield prediction of rice (b)

1949—2015 年吉林省水稻产量整体呈上升趋势, 期间由于气候因素和政策导向等导致一些年份水稻产量出现较大波动, 特别是 2005—2015 年。ARIMA 模型预测结果较好地反映了水稻产量整体上升的趋势, 但对原始数序列的波动没有做出反应。

(2) 基于 LSTM 的时间序列预测

① 建立模型

使用 Keras 深度学习库在 Python 中构建 LSTM 网络模型。参数设置为默认参数, 即

2018年10月

输入层有 1 个 input，隐藏层 4 个神经元，输出层为预测一个值，激活函数用 sigmoid，迭代 100 次，batch size 为 1。利用上面得到的平稳序列数据对模型进行训练。

②模型预测

利用得到的 LSTM 模型对 1991—2015 年水稻总产量进行预测，结果如图 9。

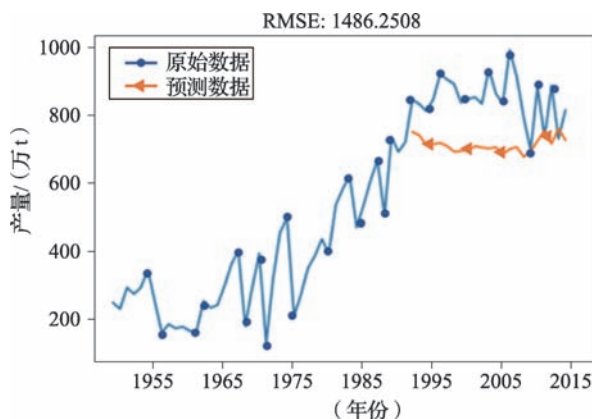


图 9 长短期记忆神经网络 (LSTM) 模型预测水稻产量

Fig.9 Prediction of rice yield using LSTM model

可以看出，LSTM 模型较预测曲线整体趋势较为平缓，预测产量普遍低于实际产量，未能反应水稻产量整体上升的总体趋势，对于原始数序列的波动反应不明显。

2.4 从横截面数据角度预测产量

实验中用于预测水稻产量的病虫害数据为吉林省历年稻瘟病发病频率（1961—2009 年）、玉米螟发生情况（1980—2007 年）、地下害虫及大豆蚜发生情况（1981—2006 年）、食心虫发病情况（1982—2006 年）、黏虫发生情况（1979—2006 年）。用于预测水稻产量的其他特征数据为水稻的竞争性作物种植面积与产量的历年数据，包括 1949—2015 年吉林省玉米、大豆种植面积、单产及总产量数据。

实验使用 sklearn 机器学习库进行数据的处理和 SVR、MLP 模型训练及预测。主要包括数据预处理、定义分类器、训练模型、模型预测等步骤。由于原始数据由多种特征数据组成，为高维数据，为了更有效处理数据，对原始数据进行主成分分析（PCA）降维处理后，再次进行上述分析流程，并与未经降维处理的预测结果进行对比。

采用每种方法建立预测模型时，考虑两种特征的组合：①考虑水稻生长受当年病虫害影响，采用病虫害数据预测当年水稻产量；②考虑竞争性作物种植面积及产量对农户种植意向的影响，采用竞争性作物种植面积及产量数据预测下一年水稻产量。两种特征组合可分别应用于当年及下一年水稻产量的预测任务。

（1）基于 SVR 的机器学习预测

对原始数据进行归一化处理后，将 1950—1990 年的数据作为训练集用于预测模型的建立，1991—2015 年的数据作为测试集用于检验模型预测效果。

使用 sklearn 机器学习库 SVR 模型对水稻产量进行预测。主要步骤为：对原始数据中的缺失值进行函数填充后，对数据进行归一化处理；选择分类器为 SVR，用训练集数

据对模型进行训练，并建立预测模型；将测试集数据于模型预测效果的检验。对原始数据进行主成分分析 PCA 降维处理后，再次进行上述分析流程。

根据以上流程，利用病虫害数据预测当年水稻产量，得到预测曲线 SVR1 和 SVR1（主成分分析降维）。利用竞争性作物种植面积及产量数据预测下一年水稻产量，得到预测曲线 SVR2 和 SVR2（主成分分析降维）（图 10）。

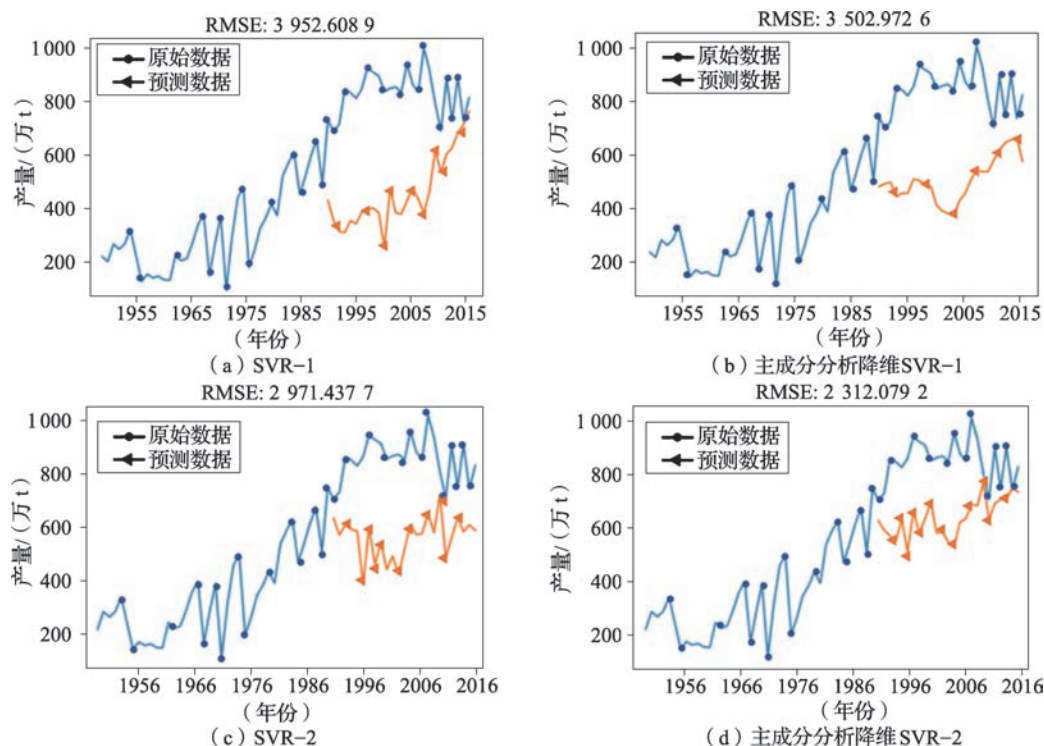


图 10 利用支持向量机（SVR）模型预测水稻产量
Fig.10 Prediction of rice yield using SVR model

SVR 模型预测曲线呈波动上升，预测产量普遍低于实际产量，在一定程度上反应了原始数序列的波动情况。由分析结果可知，采用 PCA 对影响因素进行降维处理后，SVR 模型预测准确性均有所提高。

（2）基于 MLP 神经网络的机器学习预测

使用 sklearn 机器学习库 MLP 模型对水稻产量进行预测。主要步骤为：对原始数据中的缺失值进行函数填充后，对数据进行归一化处理；选择分类器为 MLR，用训练集数据对模型进行训练，并建立预测模型；将测试集数据于模型预测效果的检验。对原始数据进行主成分分析 PCA 降维处理后，再次进行上述分析流程。

根据以上流程，利用病虫害数据预测当年水稻产量，得到预测曲线 MLP1 和 MLP1（主成分分析降维）。利用竞争性作物种植面积及产量数据预测下一年水稻产量，得到预测曲线 MLP2 和 MLP2（主成分分析降维）（图 11）。

2018 年 10 月

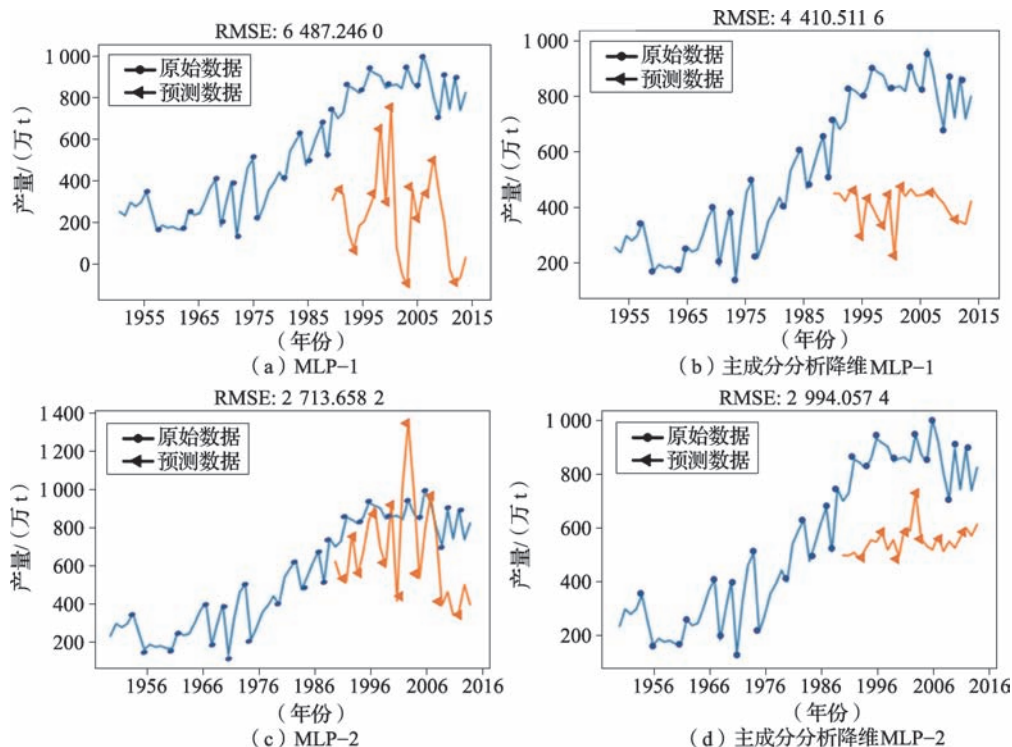


图 11 利用多层感知器 (MLP) 模型预测水稻产量

Fig.11 Prediction of rice yield using MLP model

MLP 模型预测产量普遍低于实际产量, 预测结果随原始数据序列的波动出现剧烈波动。采用 PCA 对影响因素进行降维处理后, 预测曲线波动性减缓, MLP1 模型预测准确性提高, 而 MLP2 模型预测准确性降低。

3 结果与讨论

利用吉林省病虫害和玉米大豆产量等历史数据, 基于 4 种模型对吉林省水稻产量进行预测, 预测值与实际值的 RMSE 见表 1。基于 ARIMA 模型和 LSTM 模型的时间序列预测 RMSE 相对横截面预测 RMSE 较小, 但预测值与实际值之间仍存在差距。在横截面数据预测中, 原始数据经主成分分析 PCA 降维处理后, SVR1、SVR2 和 MLP1 模型预测的 RMSE 值相对未经降维处理均有所下降, 表明在数据维度过高、数据量较小情况下, 降维处理在多数情况下可提升模型预测性能。

近年来我国对粮食生产实施了良种补贴、农机购置补贴、农民直补等一系列农业支持政策, 加之生物技术、农业工程和管理技术的发展和农田水利基础设施的不断完善, 使粮食持续增产潜力大大提升。利用 ARIMA 和 LSTM 进行时间序列预测时, 只考虑产量随时间的变化趋势, 没有考虑其他因素的变化对趋势延续性的影响, 导致对近年产量的预测与实际情况相比偏低。

表 1 不同方法预测结果 RMSE 值
Table 1 RMSE of predicting results with different methods

编号	模型	RMSE
1	自回归滑动平均模型 (ARMA)	1576.0214
2	长短期记忆神经网络 (LSTM)	1486.2508
3	支持向量机 -1 (SVR-1)	3952.6089
4	支持向量机 -1 (主成分分析降维) (SVR-1 with PCA)	3502.9726
5	支持向量机 -2 (SVR-2)	2971.4377
6	支持向量机 -2 (主成分分析降维) (SVR-2 with PCA)	2312.0792
7	多层感知器 -1 (MLP-1)	6487.2460
8	多层感知器 -1 (主成分分析降维) (MLP-1 with PCA)	4410.5116
9	多层感知器 -2 (MLP-2)	2713.6852
10	多层感知器 -2 (主成分分析降维) (MLP-2 with PCA)	2994.0574

农业生产是复杂的经济与自然再生产过程，粮食生产过程受到气象条件、作物品种、农资投入、田间管理、市场供需、国家政策等多方面的影响^[18]。而利用 SVR 和 MLP 模型进行横截面数据预测时，模型训练使用的原始数据为水稻重要病虫害发生情况、大豆种植面积与产量、玉米种植面积与产量等数据，致使模型主要反映水稻产量与病虫害发生情况及竞争性作物种植情况的关系，没有考虑其他影响产量的气象因素、投入管理因素、社会经济因素，导致模型预测产量与实际产量相比偏低。

由时间序列模型与横截面数据预测模型对比的实证分析结果可知，在进行水稻产量预测时，如缺乏影响产量的重要因素的数据，则应用时间序列模型分析结果较好。反之，如掌握影响产量的重要因素数据，则宜采用横截面数据预测模型。在进行横截面数据预测时，如分析影响因素较多，可采用 PCA 对影响因素进行降维处理，通常能够提高预测结果的准确性。此外，在进行横截面数据预测中，如训练数据较少时，采用回归模型比神经网络模型预测结果准确性高，且神经网络模型中，在训练数据量较小时，算法表现不够稳定。在进行横截面数据进行水稻产量预测时，利用竞争性作物种植面积及产量数据预测下一年水稻产量的结果优于利用病虫害数据预测当年水稻产量。

在后续研究中，拟通过补充气象数据、农资投入等数据，提高基于 SVR 和 MLP 模型的横截面数据预测准确性，并通过构建时间序列预测和横截面预测组合模型，进一步提高水稻产量预测的精度。

参考文献

- [1] 虞国平. 水稻在我国粮食安全中的战略地位分析. 新西部, 2009(11): 31~33.
- [2] Awal M, Siddique M. Rice production in bangladesh employing by Arima Model. *Bangladesh Journal of Agricultural Research*, 2011, 36(1): 51~62.
- [3] Biswas B, Dhaliwal L K, Singh S P, et al. Forecasting wheat production using ARIMA model in Punjab. *International Journal of Agricultural Sciences*, 2014, 10(1): 158~161.
- [4] Schmidhuber J, Gers F, Eck D. Learning nonregular languages: a comparison of simple recurrent networks and LSTM. *Neural Computation*, 2014, 14(9): 2039~2041.
- [5] 汤鹏杰, 王瀚漓, 许恺晟. LSTM 逐层多目标优化及多层概率融合的图像描述. 自动化学报, 2018, 44(7): 1237~1249.
- [6] 黄光许, 田垚, 康健, 等. 低资源条件下基于 i-vector 特征的 LSTM 递归神经网络语音识别系统. 计算机应用研究, 2017, 34(2): 392~396.
- [7] 周瑞, 魏正英, 张育斌, 等. 基于 LSTM 递归神经网络的番茄目标产量时间序列预测. 节水灌溉, 2018(8): 66~70.

2018年10月

- [8] Serge S, Andrea F, Jean-Noël A, et al. Crop losses due to diseases and their implications for global food production losses and food security. *Food Security*, 2012, 4(4): 519~537.
- [9] Tan Gen-jia, Dong Meng, Shen Jing-ting, et al. Disease grading criterion and assessment of yield loss caused by maize rough dwarf disease. *Plant Diseases and Pests*, 2012 (1): 1~4.
- [10] 罗正媛, 汤洋. 黑龙江省玉米产量变化的预测分析—基于支持向量机的实证研究. *农机化研究*, 2013, 35(2): 30~34.
- [11] 合尼古力·吾买尔, 林玲. Morlet 小波变换结合 MLP 神经网络的齿轮箱故障检测方法. *电子器件*, 2016, 39(4): 834~840.
- [12] 陈龙, 汪六一, 殷宪振, 等. 基于药物晶体的立体形态参数辨识硫酸氢氯吡格雷的不同晶型. *药学学报*, 2013, 48(9): 1459~1463.
- [13] 邹慧琴, 李硕, 邢姝, 等. 电子鼻技术结合 MLP 网络对不同贮藏时间西洋参的鉴别研究. *中华中医药学刊*, 2013(7): 1683~1685.
- [14] 杨铁军, 杨娜, 朱春华, 等. 一种新的基于 ARIMA 模型的粮食产量预测. *河南工业大学学报 (自然科学版)*, 2015, 36(5): 19~22.
- [15] Gers F A, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. *Neural Computation*, 2000, 12(10): 2451~2471.
- [16] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述. *电子科技大学学报*, 2011, 40(1): 2~10.
- [17] 韩敏. *人工神经网络基础*. 大连: 大连理工大学出版社, 2014.
- [18] 陈锡康, 杨翠红. 农业复杂巨系统的特点与全国粮食产量预测研究. *系统工程理论与实践*, 2002, 22(6): 108~112.

Comparative analysis of rice yield forecasting based on time series analysis and cross-sectional prediction in Jilin Province of China

Chen Wei¹, Qi Weiyan¹, Yuan Fuxiang², Li Zhemin^{1*}

(1. Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China;

2. Jilin Institute of Meteorological Sciences, Changchun 130062, China)

Abstract: [**Purpose**] Rice is one of the main food crops in China. Rice yield greatly contributes to China's food security. Rice yield forecasting is of great significance to China's agricultural development. Traditional approaches use time series or cross-sectional data, and apply statistical methods or machine learning methods to predict rice yield. However, the comparative analyses between multiple methods are still lacking. This paper aims to compare the differences between the model's forecasting performance, and to analyze the usability of models in specific circumstances, in order to better provide implications for national food security issues. [**Methods**] In this paper, we use multiple data sources, including historical rice yield, disease and pest outbreak, and other characteristics data, to forecast rice yield in Jilin province, China. Four models are used to forecast rice yield. ARIMA and LSTM are utilized for time series forecasting. Cross-sectional prediction is also conducted using SVR and MLP. The forecasting results of different models are analyzed and compared. [**Results**] Both ARIMA and LSTM have resulted in a good forecasting performance using time series data. The forecasting performance would be further improved if a principle component analysis is applied to reduce the dimensionality of the original data. [**Conclusion**] Comparing to the cross-sectional data forecasting, time series data forecasting has achieved a better performance. This indicates that a rational selection of data and model would improve the performance in rice yield forecasting.

Key words: rice yield; ARIMA; LSTM; SVR; MLP; time series