

LSTM 模型在耕地面积预测领域的构建与应用

向雁¹, 侯艳林¹, 姜文来^{2*}, 陈印军², 成良强³

1. 贵州商学院旅游管理学院, 贵阳 550014

2. 中国农业科学院农业资源与农业区划研究所, 北京 100081

3. 贵州省农业科学院油料研究所, 贵阳 550009

摘要 长短期记忆(LSTM)模型广泛应用于系统故障、交通流量、股票指数、紧急事件、碳排放、石油产量、农区地下水位等多个领域,均表现了出色的预测性能。为了丰富耕地面积预测方法、提升耕地预测精度,将LSTM模型引入耕地面积预测。选择常用的趋势外推模型、指数平滑模型、灰色模型、移动平均自回归、支持向量机、NAR动态神经网络等6类模型进行对比,并以耕地变化趋势比较复杂的黑龙江省和变化趋势比较单一的辽宁省、吉林省作为案例进行分析,以验证LSTM模型耕地面积预测效果。结果表明,从均方根误差(RMSE)、平均绝对误差(MAPE)这2个指标的综合评价来看,LSTM模型拟合和预测效果均为最优。根据LSTM模型预测,2018—2030年黑龙江、吉林、辽宁3省的耕地面积将呈持续减少的趋势,耕地减少速度均有放缓之势。

关键词 耕地;预测模型;深度学习;神经网络;LSTM模型

中国是世界第一人口大国,粮食安全事关国家安全和稳定大局。耕地是粮食生产的基础^[1],是中国最为宝贵的资源,关系十几亿人吃饭大事。当前中国经济发展进入新常态,新型工业化、城镇化建设深入推进,耕地后备资源不断减少,耕地保护面临多重压力^[2]。研究耕地面积变化规律,科学预测未

来耕地变化趋势,可为土地利用规划和耕地保护政策制定提供重要依据,对于促进耕地资源保护和保障粮食及主要农产品安全具有重要现实意义。

耕地面积变化及预测研究是学界广泛关注的热点问题。利用时间序列数据进行耕地面积预测的模型主要有趋势预测、灰色模型、移动平均模型、

收稿日期:2020-08-20;修回日期:2020-11-05

基金项目:中国农业科学院科技创新工程协同创新任务项目(CAAS-ZDRW202012)

作者简介:向雁,讲师,研究方向为农业资源管理,电子信箱:1263315496@qq.com;姜文来(通信作者),研究员,研究方向为农业资源管理与利用,电子信箱:jiangwenlai@caas.cn

引用格式:向雁,侯艳林,姜文来,等. LSTM模型在耕地面积预测领域的构建与应用[J]. 科技导报, 2021, 39(9): 100-108; doi: 10.3981/j.issn.1000-7857.2021.09.012

随机森林模型、支持向量机模型、神经网络模型,其中以趋势外推和灰色模型应用最为广泛。以上模型根据是否考虑影响因素可归纳为2类:一是基于耕地面积时间序列自身特征,利用已知数据建立拟合模型,进而利用拟合模型预测未来值。例如,陈印军等^[3]基于历史时序数据利用趋势预测模型对全国耕地面积进行了预测;张琳等^[4]利用线曲线、对数曲线和增长曲线等趋势外推预测模型预测了中国各省2015年耕地数量;季翔等^[5]利用灰色模型分别预测了洞庭湖地区2020年的旱地和水田面积。谢树春等^[6]利用指数平滑和灰色模型预测了宁夏的耕地需求量。二是基于耕地面积时间序列和相关影响因素的时间序列数据建立回归模型,进而利用相关影响因素未来值回归预测未来值。现有研究主要采取利用前序影响因素和耕地的时间序列数据,构建回归模型,然后利用后序已知影响因素数据输入回归模型,获得耕地面积预测值,并与后序耕地面积实际值比较,测试模型的预测效果,尚停留在构建模型和测试模型阶段^[7-9]。对于如何运用所建立的回归模型,利用影响因素的未来值来预测未来耕地面积变化的研究相对较少,且尚存一定的不足之处。例如,车明亮等^[10]利用相关影响因素和耕地面积数据建立的回归模型,通过研究区的官方规划文件获取各影响因素的未来值,预测了研究区的未来耕地面积,但其获取各影响因素的途径具有较大的局限性;王全喜等^[11]利用相关影响因素和耕地面积数据建立的回归模型,通过以指数平滑法分别进行预测的方式,获得各影响因素的未来值,进而预测了研究区未来耕地面积,但因未考虑各因素已知序列自身变化态势筛选适用模型,仅采用单一模型对各因素序列分别进行预测,可能会导致各因素预测误差在回归预测中累计,进而影响耕地面积预测结果的准确性和可靠性。

近年来,随着深度学习技术的发展,深度学习模型逐渐被应用到时间序列数据研究中,在众多深度学习模型中,循环神经网络(RNN)对时间序列数据分析展现出更强的适应性。作为一种特殊的循环神经网络模型,长短期记忆(LSTM)模型通过特殊的结构设计,有效弥补了常规RNN的梯度消失

和梯度爆炸、长期记忆能力不足等问题,使得循环神经网络能够有效利用长距离的时序信息^[11]。相比灰色模型、移动平均自回归(ARIMA)等线性模型和RNN、支持向量机(SVM)等非线性模型,LSTM深度学习模型对时间序列的预测性能更为出色,能够更有效地利用过去的信息预测未来时间点的信息,已经在系统故障、交通流量、股票指数、紧急事件、碳排放、石油产量、农区地下水水位等不同领域的时间序列研究中运用^[12-18]。深度学习引入耕地面积变化领域仍有待探索。

总体来看,时间序列拟合模型在耕地预测中已经得到广泛应用,但不同模型的拟合和预测精度存在差异,有必要进一步筛选精度更高的时间序列拟合模型。同时,已有研究在建立耕地面积回归模型时普遍会纳入自然、人口、社会、经济、农业发展、政策等多种影响因素,众多影响因素未来值的可获得性和准确性,严重影响了利用回归模型方法预测未来耕地面积的可操作性和可靠性。因此,在基于多因素驱动的耕地面积回归模型,其众多影响因素未来值获取也需要以高精度的时间序列拟合模型预测为基础。因此本研究的主要目的在于探求一个适用于耕地面积预测的高精度时间序列拟合模型。鉴于LSTM深度学习模型对时间序列的预测性能出色,将其引入耕地面积变化预测研究,以黑龙江省、吉林省、辽宁省耕地面积时间序列为研究对象,利用LSTM模型和对比模型分别对不同变化态势的3省耕地面积时序数据进行拟合和预测,从不同模型对同一时序数据(即同一个省区数据)的拟合及预测能力和同一模型对不同复杂度时序数据(即不同省区数据)的拟合及预测能力两个维度对所选模型进行评价,以期筛选出适应性更广、准确性更高的预测模型。

1 研究区域与数据

1.1 研究区域

东北三省包括黑龙江省、吉林省、辽宁省,均为中国粮食主产区,在保障国家粮食安全中具有举足轻重的地位。至2017年底,东北三省共有耕地面积

积 2780.40 万 hm^2 , 占全国耕地总面积的 20.61%, 人均耕地面积是全国平均水平的 2.65 倍, 粮食作物总产量 1.39 亿 t, 占全国粮食总产量的 21%。1990 年以来, 3 省耕地面积变化呈现不同变化态势, 黑龙江省耕地面积处于增减波动状态, 变化趋势较复杂且年际波动较大; 吉林省耕地面积呈先增后减变化态势, 变化趋势较复杂但年际波动幅度较小; 辽宁省耕地面积整体上呈不断下降趋势, 变化相对单一。

1.2 数据来源

耕地面积数据来源于国家统计局, 自然资源部, 及黑龙江省、吉林省、辽宁省统计局等部门公布的数据, 其中, 1990—1995 年耕地面积源自《黑龙江统计年鉴(1997)》、《吉林统计年鉴(1997)》、《辽宁省统计年鉴(1997)》; 1996 年耕地面积源自国土资源部、国家统计局、全国农业普查办公室《关于土地利用现状调查主要数据成果的公报》^[19]; 1997—2008 年耕地面积源自《中国国土资源年鉴(1999—

2009)》; 2009—2017 年耕地面积源自《中国统计年鉴(2015)》《中国统计年鉴(2018)》。

1.3 数据订正

当前, 比较系统的官方耕地面积数据为 1995 年(含)之前的国家统计局数据、基于第一次全国土地资源调查的 1996—2008 年国土资源数据、基于第二次全国土地资源调查的 2009—2017 年国土资源数据。

图 1 反映了黑龙江、吉林、辽宁 3 省耕地面积原始数据变化情况。3 省耕地面积均分别在 1995—1996 年间和 2008—2009 年间发生“突变”, 其原因是数据口径有所变化, 其中 1990—1995 年为统计口径数据, 1996—2008 年为第一次国土调查(简称“一调”)成果及变更数据, 2009—2017 年为第二次国土调查(简称“二调”)成果及变更数据。口径不同使得耕地数据系列不连续, 基于研究的需要, 有必要对数据进行订正。

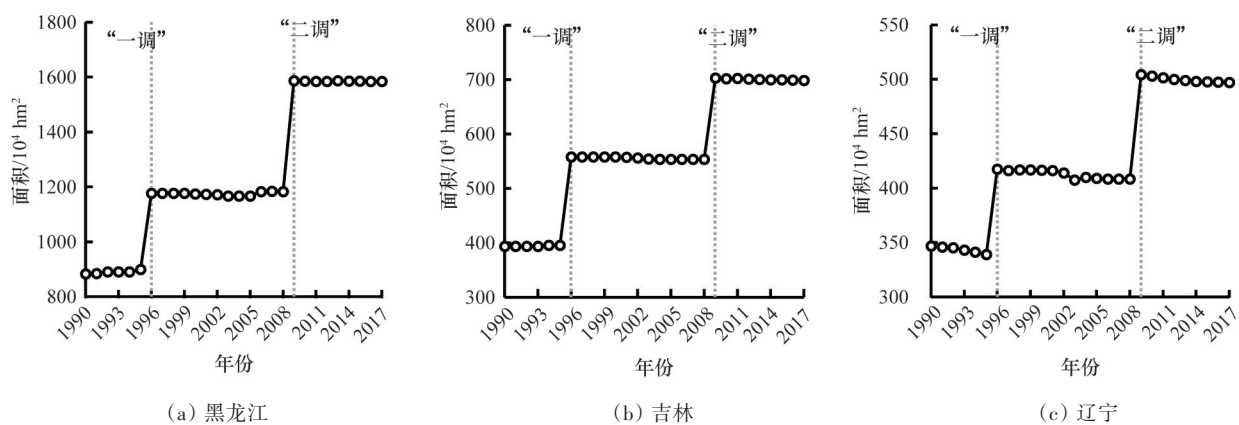


图 1 订正前耕地面积时序变化

已有研究普遍采用“基于基准数据, 结合耕地面积增减变化数据逐年反推”的思路和方法进行耕地面积数据的订正, 封志明等^[1]、汪涌等^[2]、蔡运龙等^[20]、宋小青等^[21]分别以 1996 年“一调”数据为基础, 将 1996 年之前耕地数据订正为“一调”口径数据; 陈印军等^[3]、金涛^[22]、郑亚楠等^[23]分别以 2009 年“二调”数据为基础, 将 2009 年之前耕地数据订正为“二调”口径数据。上述研究均表明, 经订正后的

数据能够更好反映耕地面积变化真实情况。本研究采取上述思路和方法进行数据订正, 即以 2009 年“二调”数据为基础, 结合耕地面积增减变化值进行逐年反推, 将 1990—2008 年 3 省耕地面积数据订正为基于“二调”口径的数据。构成基于“二调”的 1990—2017 年黑龙江、吉林、辽宁 3 省耕地面积时序数据(图 2)。经订正, 黑龙江、吉林、辽宁 3 省耕地面积数据能够更好地反映其时序变化的真实情况。

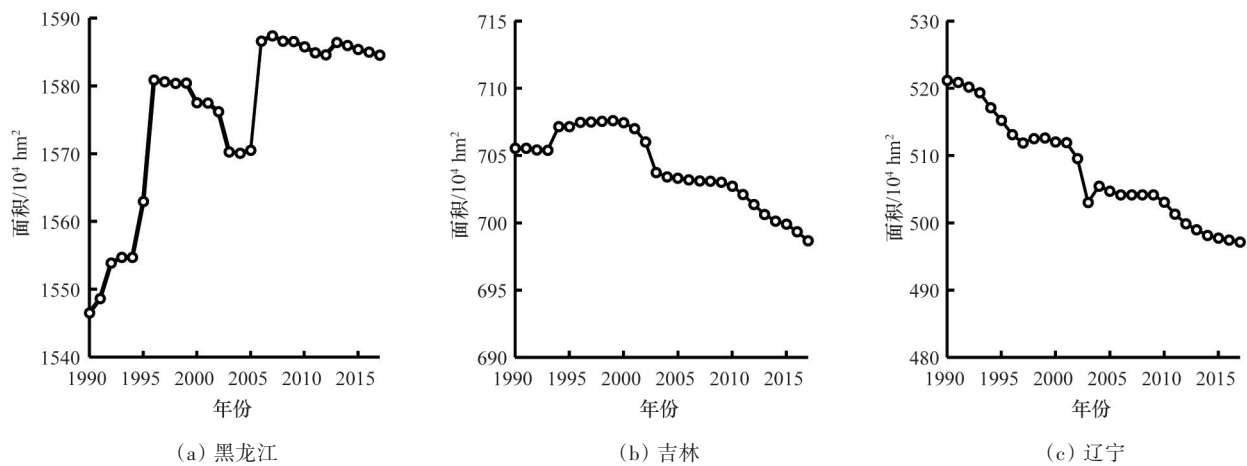


图2 订正后耕地面积时序变化

2 研究方法

2.1 划分训练和测试数据

为评价各模型对前序时间序列的拟合能力,将订正后的1990—2012年耕地面积数据作为训练集,利用LSTM模型和各对比模型分别进行拟合训练,建立拟合模型,并比较其拟合效果;为评价各模型对后序时间序列的预测能力,以2013—2017年耕地面积数据作为测试集,利用训练好的LSTM模型和各对比模型分别预测该时期耕地面积,将各模型预测值分别与实际值进行对比,评价各模型对耕地面积未来变化的预测效果。

2.2 LSTM模型

LSTM模型由Hochreiter和Schmidhuber^[24]提出,2012年由Graves^[25]改进,可以通过设计精细的网络结构来解决梯度消失问题。LSTM模型的每个神经元的结构如图3所示,其内部包括1个记忆储存(cell)和3个门控(gates)设置,cell记录神经元状态,输入门(input gate)和输出门(output gate)用来接收、输出参数和修正参数,遗忘门(forget gate)用来控制上一单元状态的被遗忘程度^[22]。

图3中, x_t 为输入值(input), h_t 为隐藏状态(hidden state), c_t 记忆单元状态(cell state), f 为遗忘门(forget gate), g 为记忆单元(memory cell), i 为输入门(input gate), o 为输出门(output gate)。

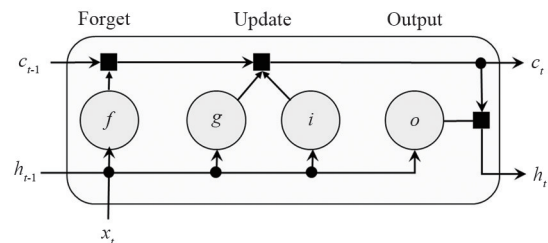


图3 LSTM单元结构

经多次训练确定模型结构及参数如下:隐含单元数量为100个的LSTM模型效果稳定,故将LSTM模型的隐含单元设为100个;采用Adam优化器,将训练的迭代次数设为1000次;梯度阈值设为1,以防止梯度爆炸;指定初始学习率为0.005,在125轮训练后通过乘以因子0.2来降低学习率。利用Matlab 2018a深度学习工具箱实现。

为了比较评估LSTM模型的拟合效果,研究选择了在耕地资源预测模型中具有代表性的趋势外推模型、指数平滑模型、灰色模型、移动平均自回归模型、支持向量机模型、NAR动态神经网络等6类模型的拟合效果进行比较。

2.3 对比模型

2.3.1 趋势外推模型(trend extrapolation, TE)

根据本研究耕地时序数据呈现的函数特性,涉及函数包括线性曲线(linear, LINE)、二次曲线(quadratic, QUADR)、三次曲线(cubic, CUB)、对数曲线(logarithmic, LOG)、逆曲线(inverse, INV)、幂

曲线(power, POW)、S曲线(S)、增长曲线(growth, GROW)等曲线拟合模型。各函数拟合及预测基于SPSS 25.0内曲线估计工具实现。

2.3.2 指数平滑模型(exponential smoothing, ES)

在实际测算中,指数平滑预测模型可以根据时间的远近赋予各时期不同的权重,使预测值具有较高的精度。指数平滑模型拟合及预测基于SPSS 25.0时间序列预测工具实现,运用了霍尔特指数平滑平衡模型(Holt's linear trend exponential smoothing model, HOLT)和布朗指数平滑模型(Brown's linear trend exponential smoothing model, BROWN)。

2.3.3 灰色模型(gray model, GM)

灰色系统理论是一种研究少数据、贫信息不确定性问题的方法,主要通过对部分已知信息的开发生成,提取有价值的信息,以实现了对系统运行的正确认识 and 有效控制。根据耕地资源和粮食生产所特有的灰色属性,结合研究内容的数据库特点,利用系统工程的思想,选取灰色系统中的基于累加方式的灰色系统GM(1,1)模型和基于累减方式的灰色系统Verhulst模型(grey Verhulst model, VERH),对耕地面积进行了模拟预测。灰色模型拟合及预测基于南京航空航天大学发行的灰色建模软件V6.0实现。

2.3.4 移动平均自回归(autoregressive integrated moving average, ARIMA)模型

ARIMA模型是时间序列预测与控制的Box-Jenkins建模方法中最为重要的基本模型,适用于非平稳时间序列的短期预测。根据检验平稳性、差分、模型定阶、参数估计、诊断检验等建模流程,对耕地面积进行了分析预测,其参数参照专家建模器推荐值及时间序列的自相关函数值(ACF)与部分自相关函数值(PACF)特征进行调整。ARIMA模型拟合及预测基于SPSS 25.0时间序列预测工具实现。

2.3.5 支持向量机(support vector machine, SVM)模型

SVM模型可以有效地实现对基于小样本的高维非线性系统精确拟合,并且采用结构风险最小原

则,具有很好的推广性。SVM模型拟合及预测基于Matlab 2018a实现,其方法为利用训练时间序列数据集,采取滑窗法构建模型训练的输入序列和目标序列进行训练拟合,建立SVM模型。利用拟合输出末期数据作为模型预测的输入序列,逐期累进预测目标序列。

2.3.6 NAR动态神经网络(NAR dynamic neural network, NARNET)

当时间序列呈现非线性特征且伴有随机性时,神经网络处理的优势就得以体现。利用Matlab 2018a机器学习工具箱中时间序列神经网络工具(neural time series tool),训练时序数据建立NAR动态神经网络,采用试错法确定拟合效果最优的神经元数量,训练获得适宜的NAR动态神经网络。

2.4 模型评价指标

选取均方根误差(RMSE)和平均绝对误差(MAPE)评价各模型预测精度优劣。

RMSE可以衡量模型拟合值与实际值之间的偏差,其取值 ≥ 0 , RMSE值越小,表明拟合值越接近于实际值,拟合效果越好;其值越大,表明拟合值越偏离于实际值,拟合效果越差。计算公式为

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

式中, n 为样本数量; y_i 为第*i*个样本实际值; \hat{y}_i 为模型对第*i*个样本的拟合值。

MAPE可以反映模型预测误差的情况。其取值 ≥ 0 ,其值越小,表明拟合值误差越小,拟合效果越好;其值越大,表明拟合值误差越大,拟合效果越差。计算公式为

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (2)$$

式中, n 为样本数量; y_i 为第*i*个样本的实际值; \hat{y}_i 为模型对第*i*个样本的拟合值。

采用RMSE、MAPE交叉检验的方式,评价模型对已知序列的拟合效果和对未来变化趋势的预测效果,当模型的RMSE和MAPE越小时,表明模型拟合效果越好。

3 模型评价与预测应用

3.1 模型拟合效果评价

为评价上述模型对前序时间序列的拟合能力,以黑龙江、吉林、辽宁3省1990—2012年耕地面积数据作为训练数据,分别利用LINE、QUADR、CUB、LOG、INV、POW、S、GROW、VERH、GM(1,1)、ARIMA、HOLT、BROWN、SVM、NARNET、LSTM建立拟合模型,计算各模型的RMSE和MAPE,结果如表1所示。不同模型对黑龙江、吉林、辽宁3省耕地面积的拟合效果显示,SVM、NARNET、LSTM对于黑龙江

省耕地面积的拟合效果优于其他模型,其中NARNET、LSTM拟合效果最优;ARIMA、HOLT、BROWN、NARNET、LSTM对于吉林省耕地面积的拟合效果优于其他模型,其中NARNET、LSTM拟合效果最优;NARNET、LSTM对于辽宁省耕地面积的拟合效果优于其他模型,其中LSTM训练拟合效果最优。总体来看,NARNET、LSTM对黑龙江、吉林、辽宁3省耕地面积的拟合效果均为最优,ARIMA、HOLT、BROWN、SVM对吉林省拟合效果较好,SVM模型对黑龙江省拟合效果较好。故选取ARIMA、HOLT、BROWN、SVM、NARNET、LSTM模型进入测试阶段。

表1 不同模型对耕地面积的训练拟合效果评价

模型	黑龙江		吉林		辽宁	
	RMSE	MAPE/%	RMSE	MAPE/%	RMSE	MAPE/%
LINE	7.54	0.40	1.29	0.16	1.67	0.25
QUADR	6.65	0.31	0.98	0.12	1.48	0.23
CUB	5.86	0.33	1.04	0.12	1.48	0.22
LOG	6.17	0.31	1.72	0.23	2.00	0.33
INV	7.70	0.43	2.03	0.26	4.34	0.80
POW	6.18	0.31	1.72	0.23	2.03	0.33
S	7.68	0.43	2.03	0.26	4.35	0.80
GROW	7.57	0.40	1.29	0.16	1.65	0.25
VERH	6.45	0.31	3.46	0.44	1.57	0.24
GM(1,1)	7.42	0.41	1.26	0.15	1.63	0.25
ARIMA	5.49	0.24	0.67*	0.06*	1.62	0.22
HOLT	5.84	0.24	0.64*	0.07*	1.64	0.23
BROWN	6.06	0.28	0.67*	0.08*	1.97	0.27
SVM	5.26	0.18**	0.75*	0.07*	1.60	0.21
NARNET	2.13**	0.07**	0.39**	0.02*	0.92**	0.11*
LSTM	0.17**	0.01**	0.13**	0.01*	0.13**	0.01**

注:*,在95%水平显著;**在99%水平显著。

3.2 模型预测效果测试

为进一步评价模型对后序时间序列的预测效果,利用上述构建效果较好的ARIMA、HOLT、BROWN、SVM、NARNET、LSTM模型分别预测黑龙江、吉林、辽宁3省2013—2017年耕地面积,分别与本省该时期实际耕地面积进行比较,并计算各模型预测的RMSE和MAPE(表2)。结果显示,对于黑龙江省和吉林省,LSTM模型预测RMSE和MAPE均为最小,预测值与实际值间误差最小,预测精度

显著高于其他模型。对于辽宁省,NARNET和LSTM模型预测的RMSE和MAPE显著小于其他模型,预测精度显著高于其他模型。NARNET模型对辽宁省预测效果较好,但对于吉林省和黑龙江省的预测效果仍弱于LSTM模型。LSTM模型能够实现黑龙江、吉林、辽宁3省耕地面积变化的高精度预测,LSTM模型较其他模型的准确性更高、适应性更广。因此选用LSTM模型对未来黑龙江、吉林、辽宁3省的耕地面积变化进行预测。

表2 不同模型对耕地面积预测的测试效果评价

模型	黑龙江		吉林		辽宁	
	RMSE	MAPE/%	RMSE	MAPE/%	RMSE	MAPE/%
ARIMA	5.31	0.27	1.14	0.15	1.20	0.19
HOLT	5.25	0.27	0.44	0.06	1.27	0.20
BROWN	2.38	0.15	0.20	0.03	2.12	0.35
SVM	3.23	0.20	0.48	0.07	0.67	0.13
NARNET	1.51	0.09	0.19	0.02	0.23*	0.04*
LSTM	0.23*	0.01*	0.12*	0.02*	0.06*	0.01*

注:*,在95%水平显著。

3.3 未来耕地面积预测结果

基于上述分析结果,LSTM模型对耕地面积变化具有良好的拟合和预测效果,可以用来预测未来耕地面积变化。为了充分利用历史时间序列信息,提高预测可靠性,利用1990—2016年黑龙江、吉

林、辽宁3省耕地面积分别训练LSTM模型。以黑龙江、吉林、辽宁3省2017年耕地面积分别进行测试,筛选预测值与实际值之间的绝对误差小于0.10万 hm^2 的LSTM模型,分别对2018—2030年黑龙江、吉林、辽宁3省耕地面积进行预测(表3)。

表3 基于LSTM模型的东北地区耕地面积预测

预测	2020年			2030年		
	黑龙江省	吉林省	辽宁省	黑龙江省	吉林省	辽宁省
PRED	1584.07	698.16	496.42	1583.36	697.6	495.52
ULC	1584.37	698.33	496.69	1584.50	698.08	496.34
LLC	1583.67	697.95	496.18	1581.97	696.90	494.70

注:PRED为预测值;ULC为90%置信区间上限;LLC为90%置信区间下限。

预测结果显示,2018—2030年黑龙江、吉林、辽宁3省的耕地面积将呈持续减少的趋势,且耕地减少速度均有放缓之势。

到2020年,黑龙江、吉林、辽宁3省耕地面积分别为1584.07万、698.16万、496.42万 hm^2 ,合计约占全国当期规划耕地保有量的22.35%。与2017年相比,2020年黑龙江、吉林、辽宁3省耕地面积将分别减少0.50万、0.51万、0.74万 hm^2 ,但仍分别高出2020年耕地保有量的规划指标196.94万、91.49万、36.29万 hm^2 ,不会突破耕地数量红线。

到2030年,黑龙江、吉林、辽宁3省耕地面积将分别为1583.36万、697.6万、495.52万 hm^2 ,合计约占全国当期规划耕地保有量的22.82%。与2017年相比,2030年黑龙江、吉林、辽宁3省耕地面积分别减少1.21万、1.07万、1.64万 hm^2 ,但在全国规划耕地保有量中比重将有所增加,预计耕地面积均可保持

在耕地数量红线范围之内。未来东北三省对于全国严守耕地红线、稳定耕地数量的支撑作用将进一步强化。

4 结论

将LSTM模型引入耕地面积预测,结果显示,LSTM模型在训练阶段和测试阶段对于黑龙江省、吉林省和辽宁省耕地面积变化的拟合效果均明显优于对比模型,尤其是在耕地面积变化趋势较为复杂的情况下表现得更为突出。相比其他对比模型,LSTM模型在耕地面积变化时间序列预测中的准确性更高、适用性更广。研究利用LSTM模型对黑龙江省、吉林省、辽宁省未来耕地面积进行预测,结果表明未来3省耕地面积将呈持续下降的趋势,但下降速度将逐渐放缓。

从同一模型对不同时间序列的拟合效果来看,各模型对于变化趋势相对简单的时间序列的拟合效果均优于变化趋势较复杂的时间序列的拟合效果。从不同模型对同一时间序列的拟合效果来看,LSTM模型和NAR动态神经网络拟合效果优于其他模型,其次为指数平滑模型、移动平均自回归模型和支持向量机模型,趋势外推模型和灰色模型的拟合效果劣于其他模型。考虑到不同模型对不同时间序列的拟合效果具有差异性,在进行时间序列预测时,应该综合考虑数据变化特征和研究需求筛选适应模型,对于变化趋势相对单一的时间序列或是对于预测精度要求较低,选取指数平滑、移动平均自回归等模型即可达到较满意的效果;对于变化趋势相对复杂的时间序列并且对于预测精度要求较高,在能获取中长期时间序列数据的情况下,建议选用LSTM模型,以降低拟合与预测误差。

随着人工智能算法研究的深入和模型的逐步修正,LSTM将在自然资源预测研究中得到广泛的应用,研究认为在LSTM模型应用上,有以下2点值得进一步探讨。

1) LSTM模型对时间序列历史数据量要求较高,在利用LSTM模型进行时间序列预测时,大多数研究通常会采用较长时期的时间序列数据进行训练拟合,以确保所建立模型的可靠性;同时,在获取长期时间序列数据难度较大的研究领域,部分研究采用了中长期时间序列进行训练拟合和预测^[26-28],也获得了较好的效果。因此,对于耕地资源、粮食生产等较难获取长时期时间序列数据的研究领域,能否基于中长期时间序列数据,利用LSTM模型进行拟合和预测,其精确性和适宜性如何,值得进一步探讨。

2) LSTM模型在部分研究领域已有较为成熟的运用,并且该模型也具备序列到序列回归拟合的能力^[16-18]。利用耕地面积和相关影响因素时间序列数据,建立LSTM回归模型对耕地面积变化的拟合和预测效果如何,能否显著优于LSTM时间序列模型的拟合和预测效果,2种预测方式孰优孰劣,也值得进一步探讨。

参考文献 (References)

- [1] 封志明, 刘宝勤, 杨艳昭. 中国耕地资源数量变化的趋势分析与数据重建: 1949~2003[J]. 自然资源学报, 2005(1): 35-43.
- [2] 汪涌, 王滨, 马仓, 等. 基于耕地面积订正的中国复种指数研究[J]. 中国土地科学, 2008, 22(12): 46-52.
- [3] 陈印军, 易小燕, 方琳娜, 等. 中国耕地资源与粮食增产潜力分析[J]. 中国农业科学, 2016, 49(6): 1117-1131.
- [4] 张琳, 张凤荣, 薛永森, 等. 中国各省耕地数量占补平衡趋势预测[J]. 资源科学, 2007, 29(6): 114-119.
- [5] 季翔, 刘黎明, 起晓星. 区域耕地粮食生产保障能力及其风险评价方法[J]. 农业工程学报, 2014, 30(7): 219-226.
- [6] 谢树春, 朱建军, 宋永永. 基于粮食安全的宁夏耕地需求量预测[J]. 农业现代化研究, 2016, 37(4): 663-670.
- [7] 张豪, 罗亦泳, 张立亭, 等. 基于遗传算法最小二乘支持向量机的耕地变化预测[J]. 农业工程学报, 2009, 25(7): 226-231.
- [8] 王霞, 王占岐, 金贵, 等. 基于核函数支持向量回归机的耕地面积预测[J]. 农业工程学报, 2014, 30(4): 204-211.
- [9] 罗亦泳, 张豪, 张立亭. 基于自适应进化相关向量机的耕地面积预测模型[J]. 农业工程学报, 2015, 31(9): 257-264.
- [10] 车明亮, 聂宜民, 刘登民, 等. 区域耕地数量变化预测方法的对比研究[J]. 中国土地科学, 2010, 24(5): 13-18.
- [11] 王全喜, 孙鹏举, 刘学录, 等. 基于随机森林算法的耕地面积预测及影响因素重要性分析——以甘肃省庆阳市为例[J]. 水土保持通报, 2018, 38(5): 341-346.
- [12] 王鑫, 吴际, 刘超, 等. 基于LSTM循环神经网络的故障时间序列预测[J]. 北京航空航天大学学报, 2018, 44(4): 772-784.
- [13] 王祥雪, 许伦辉. 基于深度学习的短时交通流预测研究[J]. 交通运输系统工程与信息, 2018, 18(1): 81-88.
- [14] 杨青, 王晨蔚. 基于深度学习LSTM神经网络的全球股票指数预测研究[J]. 统计研究, 2019, 36(3): 65-77.
- [15] Cortez B, Carrera B, Kim Y J, et al. An architecture for emergency event prediction using LSTM recurrent neural networks[J]. Expert Systems with Applications, 2018, 97: 315-324.
- [16] Huang Y S, Shen L, Liu H. Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China

- [J]. Journal of Cleaner Production, 2019, 209: 415–423.
- [17] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. Neurocomputing, 2018, 323: 203–213.
- [18] Zhanga J, Zhub Y, Zhanga X, et al. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas[J]. Journal of Hydrology, 2018, doi: 10.1016/j.jhydrol.2018.04.065.
- [19] 国土资源部, 国家统计局, 全国农业普查办公室. 关于土地利用现状调查主要数据成果的公报[J]. 中国统计, 1999(12): 5, 14.
- [20] 蔡运龙, 汪涌, 李玉平. 中国耕地供需变化规律研究[J]. 中国土地科学, 2009, 23(3): 11–18, 31.
- [21] 宋小青, 吴志峰, 欧阳竹. 1949年以来中国耕地功能变化[J]. 地理学报, 2014, 69(4): 435–447.
- [22] 金涛. 中国粮食作物种植结构调整及其水土资源利用效应[J]. 自然资源学报, 2019, 34(1): 14–25.
- [23] 郑亚楠, 张凤荣, 谢臻, 等. 中国粮食生产时空演变规律与耕地可持续利用研究[J]. 世界地理研究, 2019, 28(6): 120–131.
- [24] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735–1780.
- [25] Graves A. Supervised Sequence Labelling with Recurrent Neural Networks[J]. Berlin Heidelberg: Springer, 2012: 5–13.
- [26] 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[J]. 节水灌溉, 2018(8): 66–70.
- [27] 陈威, 祁伟彦, 袁福香, 等. 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[J]. 中国农业信息, 2018, 30(5): 95–105.
- [28] 刘炳春, 齐鑫, 王庆山. 北京城市代谢预测研究——基于长短期记忆神经网络模型[J]. 地理科学进展, 2019, 38(6): 851–860.

Establishment and application of LSTM model for cultivated land area prediction

XIANG Yan¹, HOU Yanlin¹, JIANG Wenlai^{2*}, CHEN Yinjun², CHENG Liangqiang³

1. Tourism Management School, Guizhou University of Commerce, Guiyang 550014, China

2. Institute of Agricultural Resources and Agricultural Regionalization, Chinese Academy of Agricultural Sciences, Beijing 100081, China

3. Oil Research Institute, Guizhou Academy of Agricultural Sciences, Guiyang 550009, China

Abstract The long-short term memory model (LSTM) is a special recurrent neural network structure, which is widely used in system failure, traffic flow, stock index, emergency event, carbon emission, water table depth, and other fields, showing excellent prediction performance. This paper introduces the LSTM model into forecasting cultivated land area to enrich predicting methods and improve prediction accuracy. To verify the validity of the LSTM model in cultivated land area prediction, TE, GM, ES, ARIMA, SVM and NARNET models are selected for comparison, in which Heilongjiang, Jilin and Liaoning provinces are taken as case areas for revealing evaluation effects of different time series models. The results indicate that the prediction effect of LSTM is better than other models in terms of the comprehensive evaluation of RMSE and MAPE. Finally, according to LSTM forecast, the cultivated land areas of Heilongjiang, Jilin and Liaoning provinces will continue to decrease from 2018 to 2030 and the decrease rate will slow down.

Keywords cultivated land; forecast model; deep learning; neural network; LSTM ●



(责任编辑 刘志远)