

基于山东省不同模型的物流需求预测比较研究

徐晓燕^a, 杨慧敏^a, 吕修凯^b, 王雪^a, 康静彩^a

(东北林业大学 a.工程技术学院 b.机电工程学院, 哈尔滨 150000)

摘要: **目的** 过对不同预测方法的误差对比研究, 选取预测生鲜农产品物流需求量更精准方法, 为疫情情况下山东省生鲜农产品市场进行科学性、合理化决策提供参考。**方法** 公路货物周转量、互联网普及率、GDP、人口数量、第一产业增加值等十大影响因素作为自变量, 以生鲜农产品的需求量作为因变量, 分别将小波神经网络、人工神经网络 (BP)、遗传算法优化神经网络 (GA-BP)、粒子群优化神经网络 (PSO-BP)、长短时记忆网络 (LSTM) 等 5 种方法的数据预测进行对比分析。**结果** 波神经网络和 BP 神经网络的预测值明显低于真实值, 且平均相对误差接近 20%, 而优化后的 GA-BP、PSO-BP、LSTM 算法误差均小于 5%, 分别为 4.06%、1.162%、0.45%, 因此, LSTM 预测精度最高, 效果最好。**结论** 来山东省的生鲜农产品需求量将持续增长, LSTM 算法以其精确度更高, 学习能力更强的优点, 将会被更多地应用到物流领域研究中。

关键词: 小波神经网络; 人工神经网络; 遗传算法优化神经网络; 粒子群优化神经网络; 长短时记忆网络; 需求预测

中图分类号: TB485.3 文献标识码: A 文章编号: 1001-3563(2022)23-0207-09

DOI: 10.19554/j.cnki.1001-3563.2022.23.025

Comparative Research on Forecast of Logistics Demand in Shandong Province Based on Different Models

XU Xiao-yan^a, YANG Hui-min^a, LYU Xiu-kai^b, WANG Xue^a, KANG Jing-cai^a

(a.Technology College b.Electromechanical Engineering College, Northeast Forestry University, Harbin 150000, China)

ABSTRACT: The work aims to compare and study the errors of different prediction methods to select a more accurate method for predicting the logistics demand of fresh agricultural products, and provide a reference for scientific and rational decision-making in the fresh agricultural product market in Shandong Province under the epidemic situation. With ten influencing factors, such as highway cargo turnover, Internet penetration rate, GDP, population, and added value of the primary industry, as independent variables, and the demand for fresh agricultural products as the dependent variable, the data prediction of five methods such as wavelet neural network, BP neural network, BP neural network by genetic algorithm (GA-BP), BP neural network by particle swarm (PSO-BP), long short-term memory (LSTM) were compared and analyzed. The predicted values of wavelet neural network and BP neural network were obviously lower than the actual values, and the average relative error was close to 20%, while the errors of optimized GA-BP, PSO-BP and LSTM algorithms were all less than 5%, which were 4.06%, 1.162% and 0.45% respectively. Therefore, LSTM had the highest pre-

收稿日期: 2021-12-29

基金项目: 中央高校业务经费 (2572016CB11); 校级教育教学研究项目 (DGY2020-42)

作者简介: 徐晓燕 (1997—), 女, 东北林业大学硕士生, 主要研究方向为冷链物流需求预测研究。

通信作者: 杨慧敏 (1980—), 女, 博士, 副教授, 主要研究方向为智慧物流。

diction accuracy and the best effect. In the future, the demand for fresh agricultural products in Shandong Province will continue to grow, and the LSTM algorithm will be more applied in the field of logistics research due to its advantages of higher accuracy and stronger learning ability.

KEY WORDS: wavelet neural network; BP neural network; BP neural network by genetic algorithm; BP neural network by particle swarm; long short-term memory; demand forecast

随着人们生活水平的不断提高,生鲜农产品在生活中的比重也越来越大,对生鲜农产品的安全和保鲜的关注度也逐渐提高。山东省作为农业大省,虽然生鲜农产品资源丰富,农业增加值连续多年位居全国第一,但是随着生产端和流通环节的冷链需求量不断扩大,一些冷链物流问题逐渐出现:例如生鲜农产品的损失、浪费,物流冷链运输的成本增加的现象等。如何降低生鲜农产品冷链物流成本、缩小物流企业资源的浪费程度、满足消费者对生鲜农产品更高标准的日常需要迫在眉睫。在此基础上,突如其来的新冠疫情对生鲜物流需求造成了巨大影响,通过对物流需求进行更加精准的预测,有助于促进物流资源的合理配置,实现产品供求平衡,尽可能减少新冠疫情对冷链物流需求带来的负面影响,对于生鲜农产品的物流规划以及充分实现物流系统的各项功能具有较强的现实意义。

近几年,国内的文章主要集中在对灰色预测模型及智能优化算法的单个模型研究分析,而国外则比较多关注在大数据发展下人工智能方法及其组合改进模型的应用,且对于物流供应链中企业库存、备件的资源优化配置方面研究较多^[1]。例如邓磊等^[2]采用传统预测模型即灰色预测模型,对物流需求量进行了预测。孟庆龙等^[3]使用多元线性回归模型对苹果可溶性固形物含量进行了预测分析。杨玮等^[4]使用 Montle Carlo 季节指数平滑和作业产能模型,使订单预测结果精度提高了 45%。而国外学者已经着手对深度学习预测方法的研究,例如 Ham 等^[5]训练了一个关于历史模拟的卷积神经网络(CNN),验证了 CNN 模型指数的全季节相关能力远高于目前最先进的动态预报系统。Kim 等^[6]针对传染病流行难以预测性,采用支持向量机(SVM)、半监督学习(SSL)和深度神经网络(DNN)进行预测,以检测新发传染病模式。Munkhdalai 等^[7]采用新型深度学习模型(MLP-SUL)对邮政物流需求量做了预测,发现了深度学习模型优于邮件中心数据集的最新基准。Werner 等^[8-9]还通过对新混合模型和神经网络的原有性能进行改进,得出了更加精确的预测分析。

与此同时国内学者也开始了对组合改进模型的探索,例如李德奎等^[10]以上海地铁数日数据为例,建立了 ARIMA 与 LSTM 模型,将 2 种方法对比,证明了 LSTM 方法的均方根误差相对较小。刘蓉晖等^[11]将基于改进 K-means 聚类算法的 ARMA 预测模型、单纯的 ARMA

预测法及考虑气象因素的传统 K-means 聚类的 ARMA 预测法进行预测结果分析,证明了改进的模型预测能够提高短期负荷预测的精度。马佳玉等^[12]、黄凯等^[13]通过组合模型计算,提高了预测精度。

虽然模型的组合、改进及对比已经开始被引用到数据预测分析方向,但是物流领域的需求预测研究,仍然较多的使用单一的传统预测方法比如灰色预测模型,进行简单地组合及改进分析。因此本文将能进行预测的五种模型分析引用到对生鲜农产品的需求预测,通过对比,选择精度高预测模型,对未来五年需求量的实施精准预测。考虑到当下疫情的影响使生鲜电商的发展极大的影响了冷链物流的发展模式与发展进程,文中将生鲜电商的发展(互联网普及率)考虑在影响因素里,能对山东省生鲜农产品冷链物流的现状 & 发展进行更贴合时代背景的分析 and 预测。

1 模型构建

1.1 小波神经网络模型

小波神经网络是小波和神经网络的有机融合,它继承了人工神经网络的近似能力的优点,同时具备了小波更强的学习能力,在时间预测领域得到了广大工程师的信赖^[14]。小波神经网络结构见图 1^[15]

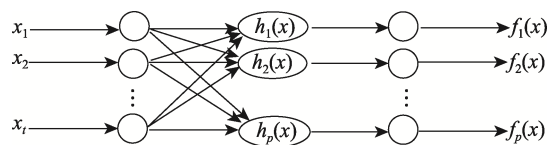


图 1 小波神经网络结构

Fig.1 Diagram of wavelet neural network structure

隐含层的激励函数可用式(1)表示。

$$h_o(x) = \prod_{j=1}^l g\left(\frac{x_j - b_{jo}}{a_{jo}}\right) \quad (1)$$

式中: j 为网络中不同的输入小波元; o 为网络中间层代码; g 为相应的小波运算; x 为网络输入量; a_{jo} 和 b_{jo} 为变换后新的伸缩和平移参数。

输出函数可用式(2)表示。

$$f_j(x) = \sum_{o=1}^p w_{jo} h_o(x) \quad (2)$$

式中: p 为小波网络的层级数; w_{jo} 为输出权重。

1.2 BP 神经网络模型

BP 神经网络包括输入层、隐含层和输出层 3 层结构, 与其他数学模型相比, BP 神经网络有着特殊的拓扑结构, 使其拥有高效的适应能力、较好的自我学习能力以及强大的非线性映射能力。

BP 神经网络预测生鲜产品物流需求量见图 2, 输入层神经元个数 $n=10$, 输出层神经元个数 $n=1$, 以 10 个影响因素作为输入层, 以物流需求量作为输出层。通常需要结合研究的生鲜农产品的物流需求量的神经网络结构, 研究隐含层神经元数量对神经网络预测精度的影响, 找到最优的隐含层神经元数, 以此改进神经网络算法。

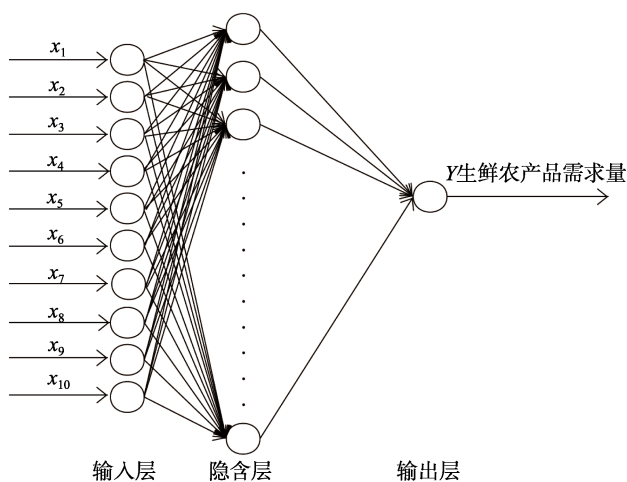


图 2 BP 神经网络结构

Fig.2 Diagram of BP neural network structure

H 表示输入层、 I 表示隐含层、 G 表示输出层, h 、 i 、 g 分别用表示对应的神经元, W_{hi} 表示输入层与隐含层权值, W_{ig} 表示隐含层与输出层权值, n 表示神经元输入, u 表示激励输出。假设训练样本集 $X=[X_1, X_2, \dots, X_N]$ 中任一训练样本 $X_m=[x_1, x_2]$ 对应的预测值为 y_m , 期望值为 z_m , 则有:

$$n_i^I = \sum_{h=1}^H \sum_{a=1}^2 W_{hi} x_{ha}, u_i^I = f(n_i^I) \quad (3)$$

$$n_g^G = \sum_{i=1}^I W_{ig} n_i^I, u_g^G = f(n_g^G) \quad (4)$$

$$y_m = u_g^G, (H=2, I=5, G=1) \quad (5)$$

1.3 GA-BP 算法模型

GA-BP 算法即利用遗传算法优化的 BP 神经网络, 遗传算法是一种全局随机搜索的采用模拟生物进化过程的遗传机理来得到最优解的智能算法^[16]。该算法利用遗传算法的全局搜索能力来寻找最优的适应度值并赋值给 BP 神经网络, 解决了 BP 神经网络收敛速度慢且易陷入局部极限值等问题。

文中收集大量物流量的影响因素数据, 根据时间先后进行整理排列, 得到物流量各项影响指标。首先

以遗传算法产生的连接权值和阈值作为 BP 神经网络的初始种群, BP 神经网络的输出误差作为适应度函数, 根据适应度函数对每个个体进行评价, 选择部分个体进行交叉和变异, 形成新的种群。然后进行进化代数比较从而选出最优个体, 并对最优个体进行解码, 得到最优 BP 神经网络的连接权值和阈值。根据最优的连接权值和阈值以及其他参数对生鲜农产品物流历史数据进行学习, 通过 BP 神经网络构建物流量预测模型。最后通过模型对未来的物流量变化趋势进行分析, 得到相应的预测结果。

1.4 PSO-BP 算法模型

PSO-BP 算法即改进粒子群优化的神经网络算法, 为解决标准粒子群不足的问题, 借助遗传算法的变异思想, 在粒子群算法中增加了 2 种因子^[17]。PSO 算法利用实数编码, 具有比遗传算法的运行速度快的优点。其不仅能寻找最优权值和阈值, 从而解决用传统 BP 神经网络单纯利用误差反向传播来调整权值和阈值容易陷入局部最优解的问题, 还可以通过群体信息共享和自身历史位置更新搜索方向, 方法参数简单、易于实现。

1.5 LSTM 算法模型

RNN 在其基础上可对时间变化数据进行建模, 但因为节点无法对比较远的节点信息进行记忆, 所以会发生梯度消失和梯度爆炸现象。LSTM 是一种特殊的 RNN, 记忆单元中存在输入门、输出门、遗忘门, 记忆单元的遗忘门能筛选信息, 有用的信息被留下, 传递到下一时刻, 无用的信息被遗忘, 有效避免梯度爆炸现象^[18]。主要结构算法如下。

$$g_x = \sigma(V_g \cdot [k_{x-1}, t_x] + b_g) \quad (6)$$

$$h_x = \sigma(V_h \cdot [k_{x-1}, t_x] + b_h) \quad (7)$$

$$\tilde{w}_x = \tan k(V_w \cdot [k_{x-1}, t_x] + b_w) \quad (8)$$

$$w_x = g_x \cdot w_{x-1} + h_x \cdot \tilde{w}_x \quad (9)$$

$$o_x = \sigma(V_o \cdot [k_{x-1}, t_x] + b_o) \quad (10)$$

$$k_x = o_x \cdot \tan k(w_x) \quad (11)$$

式中: g_x 为遗忘门; h_x 为输入门; $\tilde{w}_x$ 为当前输入单元状态; w_x 为当前时刻单元状态; o_x 为输出; k_x 为最终输出单元的值; V 为各门对应的权重矩阵; b 为偏项; $[k_{x-1}, t_x]$ 为 2 个向量连接; σ 将实数映射到 $[0,1]$ 的 sigmoid 函数; $\tan k$ 将实数映射到 $[-1,1]$ 的双曲正切函数^[19]。

2 生鲜农产品需求量影响因素分析

2.1 构建生鲜农产品市场需求指标体系

构建预测指标体系是进行需求预测的基础, 文中从研究山东省的物流经济现状出发, 结合实际性、重要性、

科学性的原则,将影响生鲜农产品市场需求量(y)的因素归结为人口规模、互联网发展、物流规模、经济规模、消费水平、产业结构 6 大类。通过借鉴其他相关论文选取经验,考虑到对于物流的影响程度的问题,通过查阅山东省统计公报及年鉴提取了十个关键指标数据,分别为公路货物周转量(X_1 ,单位:亿吨)、互联网普及率(X_2 ,单位:%)、GDP(X_3 ,单位:亿元)、人口数量(X_4 ,单位:万人)、第一产业增加值(X_5 ,单位:

亿元)、第三产业比重(X_6 ,单位:%)、城镇居民人均可支配收入(X_7 ,单位:元)、公路运输量(X_8 ,单位:亿吨)、固定资产投资增长率(X_9 ,单位:%)、物流费用占 GDP 比重(X_{10} ,单位:%),见表 1。

2.2 数据收集与整理

通过对山东省统计局和年鉴的数据进行收集整理,得到了 2001 年至 2020 年的历史数据,见表 2。

表 1 生鲜农产品物流需求指标体系
Tab.1 Logistics demand index system for fresh agricultural products

一级指标	二级指标
人口规模	人口数量
互联网发展	互联网普及率
	公路运输量
物流规模	公路货物周转量
	固定资产投资增长率
经济规模	GDP
消费水平	物流费用占 GDP 比重
	城镇居民人均可支配收入
产业结构	第一产业增加值
	第三产业比重

表 2 2001—2020 年山东省生鲜农产品冷链物流需求量及影响指标数据
Tab.2 2001-2020 cold chain logistics demand and impact index data of fresh agricultural products in Shandong Province

年份	生鲜农产品 需求量 Y / 万吨	公路货物 周转量 X_1 / 亿吨	互联网普 及率 X_2 /%	GDP X_3 / 亿元	山东省常住 人口数量 X_4 / 万人	第一产业 增加值 X_5 / 亿元	第三产业 比重 X_6 /%	城镇居民人 均可支配收 入 X_7 /元	公路运输 量 X_8 /亿 吨	固定资 产增长率 X_9 /%	物流费用占 GDP 比重 X_{10} /%
2001	10 000.98	429	2.6	9 438.3	9 041	1 359.5	36.3	7 101.1	8.4	9.3	21.13
2002	11 004.51	477	4.6	10 552.06	9 082.3	1 390	36.5	7 614.5	9.2	25.1	21.47
2003	11 247.59	527.6	6.2	12 430	9 125	1 505	34.4	7 418.4	9.8	51.7	21.4
2004	12 130.14	596.1	7.2	15 490.7	9 180	1 778.3	33.6	9 437.8	10.7	37.7	21.32
2005	12 232.67	711.8	8.5	18 468.3	9 248	1 927.6	32.1	10 744.8	12	38.9	18.57
2006	12 449.39	845.1	10.4	21 846.7	9 309	2 138.9	32	12 192	13.7	19.6	18.34
2007	12 694.12	1 069.3	16	25 887.7	9 367	2 509.1	33.2	14 265	16.4	24.2	18.41
2008	13 050.72	1 369.9	22.6	31 972.1	9 417.2	3 002.7	33.4	16 305	19.8	23.1	18.14
2009	13 796.62	6 045	28.9	33 805.3	9 470.3	3 226.6	34.1	17 811	25.2	23.3	18.11
2010	13 496.73	6 216.8	34.3	39 416.2	9 579	3 588.3	36.6	19 946	26.4	22.3	17.83
2011	13 568.47	6 624.4	38.3	45 429.2	9 637.3	3 973.8	38.3	22 792	27.9	21.8	17.8
2012	13 750.92	7 059.2	42.1	50 013.2	9 684.87	4 281.7	40	25 755	29.7	20.2	17.7
2013	14 034.80	7 498.9	45.8	54 684.3	9 733.39	4 742.6	41.2	28 264	31.2	19.6	17.6
2014	14 257.04	5 711.4	47.9	59 426.6	9 789.43	4 798.4	43.5	29 222	23	15.8	16.2
2015	14 467.10	5 738.6	50.3	63 002.3	9 847.16	4 979.1	45.3	31 545	24.2	13.9	16
2016	14 522.75	6 071.4	53.2	67 008.2	9 946.64	4 929.1	47.3	34 012	25	10.5	14.8
2017	14 598.87	6 650.2	55.8	72 678.2	10 005.83	4 876.7	48	36 789	28.8	7.3	14.2
2018	14 623.42	6 859.7	57.7	76 469.7	10 047.24	4 950.5	49.5	39 549	31.3	4.1	14.15
2019	14 452.24	7 085	64.5	71 067.5	10 070.21	5 116.4	52.8	42 329	32.6	3.8	14.7
2020	14 678.72	6 784.4	70.4	73 129	10 152.75	5 363.8	53.6	43 726	26.7	3.6	14.5

数据来源: 1.山东省统计年鉴(2001—2020); 2.山东省统计公报(2001—2020); 3.中国互联网发展状况统计公报; 4.中国互联网络发展状况统计报告。

2.3 基于灰色关联度的影响因素分析

利用灰色关联度分析法进行影响因素数据关联度分析,灰色关联分析法是用于计算评分值与“参考值”的关联程度,进而针对评价项进行评价。首先确定参考值,数据进行无量纲化处理;然后计算得出关联系数,即关联系数表示影响因素与需求量的相关程度,见图 3;最后结合关联系数值计算关联度。

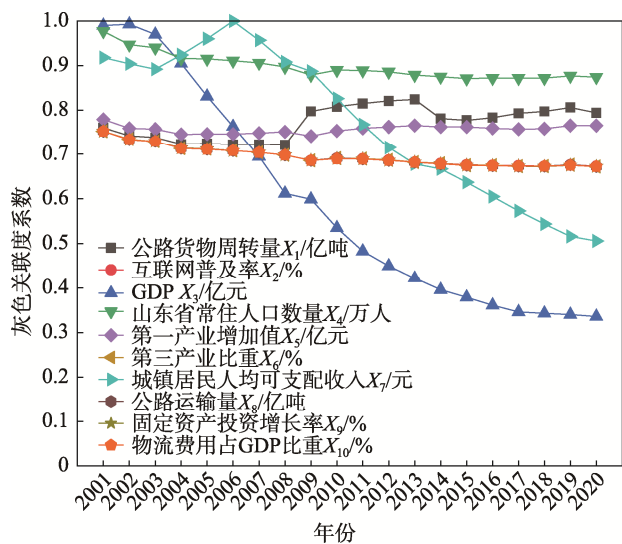


图 3 灰色关联系数图
Fig.3 Diagram of grey correlation coefficient

结合上述关联系数结果进行加权处理,得到关联度值,使用关联度值对 20 个评价对象进行评价排序;关联度值介于 0—1,该值越大表示其与“参考值”(需求量)之间的相关性越强,也即意味着评价越高。如表 3 所示,10 个评价项中,山东省常住人口数量 X_4 的综合评价最高为 0.902,各关联度都在 0.6 以上,可见影响因素选取合理,可信度较高。

3 基于 LSTM 模型山东省生鲜农产品物流需求预测分析

以山东省 2001—2020 年的山东生鲜农产品物流需求量作为目标变量,影响因子构建的指标体系作为目标预测的属性变量,通过模型计算,将预测结果误差进行对比实验。实验中统一使用 2001—2015 年的需求量及对应的影响因子作为训练集,2016—2020 年的需求量及影响因子作为测试集,通过相对误差分析评价其预测精度。

3.1 算法的参数选择

计算过程中发现,通过模型直接进行预测可能会发生过拟合现象^[20]。过拟合是指为了得到一致假设而使假设变得过度严格,导致在后期测试的时候不能很

好地识别数据,通过设置不同的迭代次数,得出了迭代次数与平均误差变化图,可以清晰观察是否发生过拟合现象,从而对模型参数进行调整,能尽量提高预测精度^[21]。如图 4 所示,BP 模型的迭代次数不超过 10 000 次时,预测误差与测试误差都在减小。之后训练误差继续减少并趋于稳定,而测试误差开始增大,由此可以得出预测过拟合了,因此设置迭代次数为 70 000,平均误差趋于稳定且精度较好。PSO-BP 模型的迭代次数不超过 100 次时,预测误差与测试误差都在减小。之后训练误差继续减少并趋于稳定,而测试误差则开始缓慢增大,由此存在轻微过拟合现象,当设置迭代次数为 100 时,训练误差尽量小的同时,实现测试集数据误差最小,见图 5。LSTM 模型的预测误差与测试误差都在随着迭代次数的增加而减小并逐渐稳定,见图 6,由此可以得出预测时没有出现过拟合现象。

表 3 各指标影响因素与需求量的关联度
Tab.3 Correlation between influencing factors of each index and demand

评价项	关联度
公路货物周转量 X_1 /亿吨	0.782
互联网普及率 X_2 /%	0.707
GDP X_3 /亿元	0.607
常住人口数量 X_4 /万人	0.902
第一产业增加值 X_5 /亿元	0.767
第三产业比重 X_6 /%	0.707
城镇居民人均可支配收入 X_7 /元	0.778
公路运输量 X_8 /亿 t	0.707
固定资产投资增长率 X_9 /%	0.707
物流费用占 GDP 比重 X_{10} /%	0.707

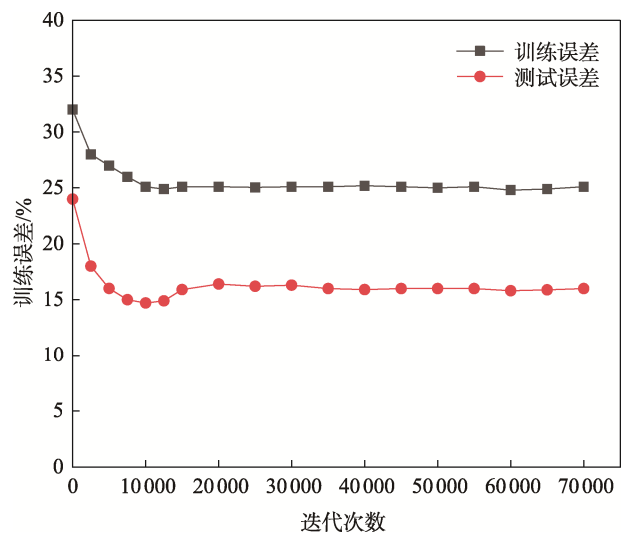


图 4 BP 神经网络误差
Fig.4 Error of BP neural network

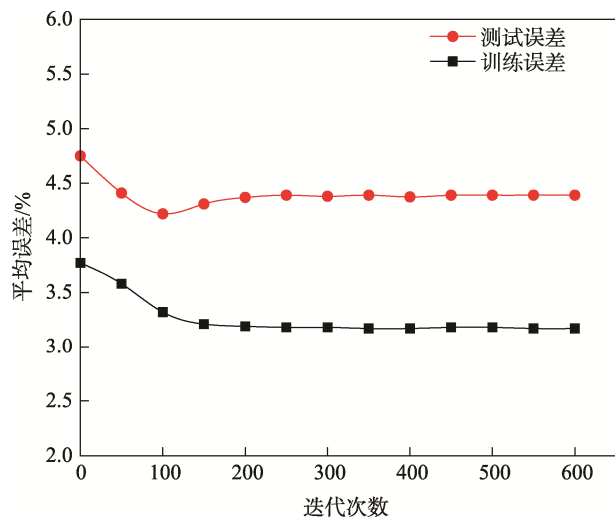


图 5 BP 优化模型误差
Fig.5 Error of BP optimization model

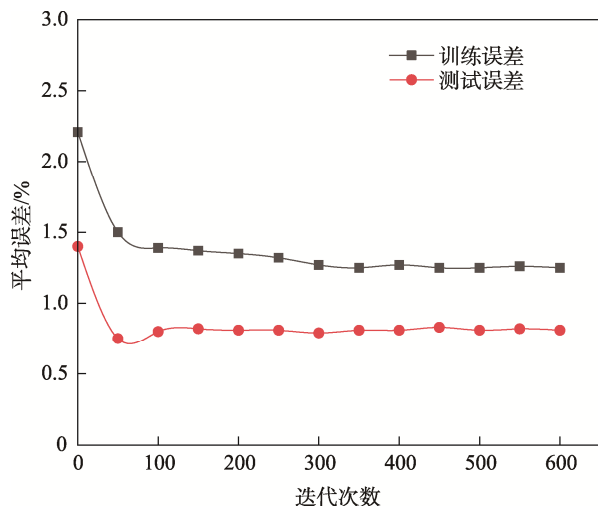


图 6 LSTM 模型误差
Fig.6 LSTM model error

通过调试分析,在误差相对稳定且小的情况下,对各个模型的进行参数设置,小波神经网络的隐形节点个数为 5, $R_1=0.01$ 、 $R_2=0.001$,迭代次数为 100。设置 BP 神经网络的输入层为 10 个神经元,输出层为 1 个神经元,训练次数为 70 000,动量系数为 0.8,学习率为 0.05,收敛误差为 0.001。BP 神经网络的隐含层通过式 (12) 确定,最终隐含层神经元数为 10 个,经验公式为:

$$a = \sqrt{m+n} + l \quad (12)$$

式中: a 为隐含层神经元数量; m 为输入层神经元数量; n 为输出层神经元数量; l 为随机选取 1—10 的常数。粒子群算法优化 BP 神经网络设置中,粒子群的粒子数为 30, $C_1=C_2=2$,迭代次数为 100 次,期望误差最小值为 0.000 1,权重 $w=0.2$ 。设置遗传算法

优化 BP 神经网络的系数,交叉概率为 0.8,变异概率为 0.1,种群规模为 15,迭代次数为 100。LSTM 模型设置迭代次数 250,梯度阈值为 1,学习率为 0.005,进行 200 次训练,得出结果见图 7,即对不同的隐层单元预测效果进行评估,不同的隐层单元的平均绝对误差不同。

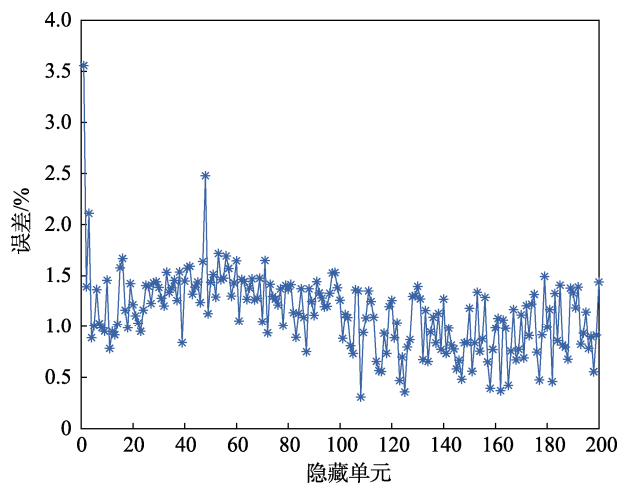


图 7 不同隐层单元预测精度的比较
Fig.7 Comparison of prediction accuracy of different hidden layer units

3.2 预测模型分析与对比

由图 8—11 可知, BP 神经网络模型的拟合情况一般,显然优化后的算法预测值曲线更加靠近实际值,预测值拟合效果较好。而基于粒子群算法的预测结果对比遗传算法预测值曲线更靠近真实值曲线,精度有所提高。LSTM 模型较其他 3 种方法来看明显更优,精度相对更好。

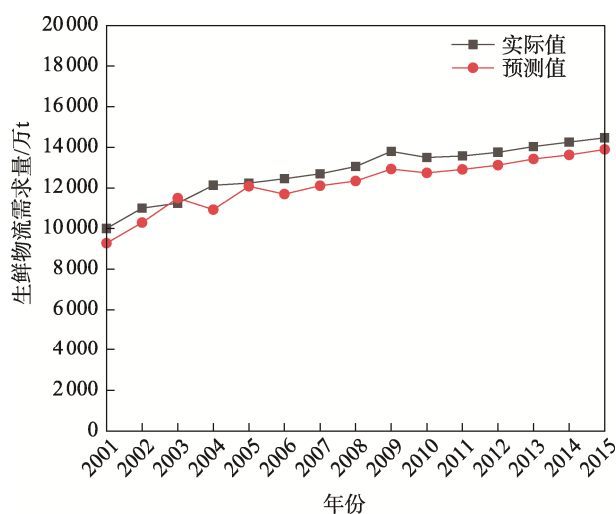


图 8 BP 神经网络模型预测值与实际值的对比
Fig.8 Comparison of BP neural network model predicted value and actual value

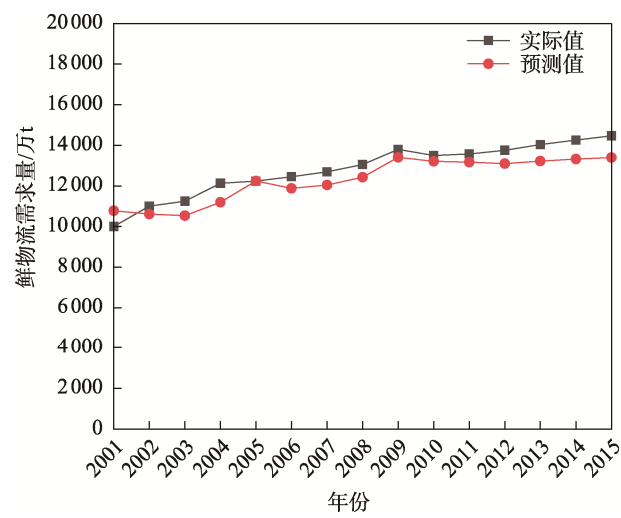


图 9 GA-BP 模型预测值与实际值的对比
Fig.9 Comparison of GA-BP model predicted value and actual value

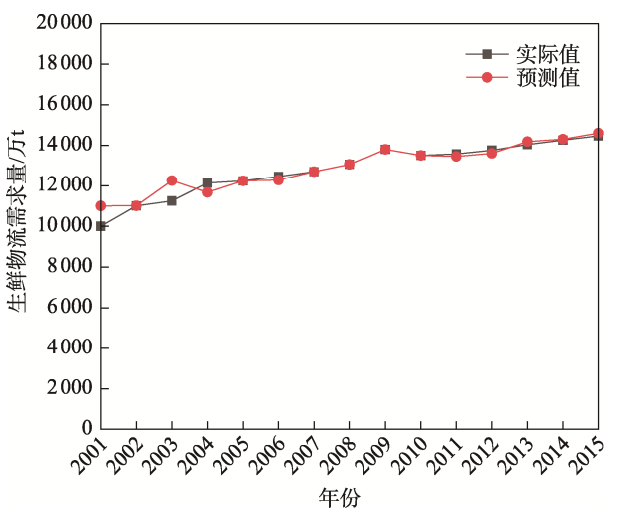


图 11 LSTM 模型预测值与实际值的对比
Fig.11 Comparison of LSTM model predicted value and actual value

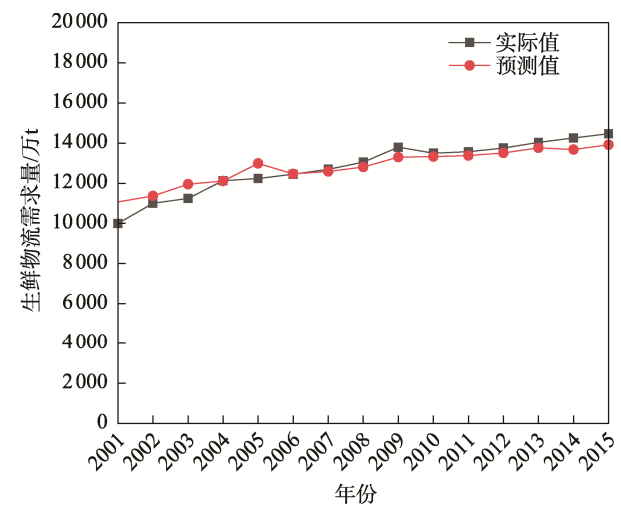


图 10 PSO-BP 模型预测值与实际值的对比
Fig.10 Comparison of PSO-BP model predicted value and actual value

由表 4 可知, 为了提高评估预测的准确性, 采用训练集的均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对百分误差 (MAPE) 3 个指标对结果进行对比分析。可以看出优化后的 GA-BP 相对

于 BP 的平均绝对百分误差降低了 19.7%, PSO-BP 相对 BP 降低了 22.3%, 因此改进后的神经网络算法的明显提高了预测精度。改进后的 PSO-BP 的均方根误差比 GA-BP 减小了 63.9%, 平均绝对误差减小了 67.8%, 绝对百分误差降低了 2.6%, 因此基于物流需求量的预测的粒子群优化神经网络算法明显更优。LSTM 与 PSO-BP 相比, 虽然 3 种指标差距不大, 但其误差都低于 PSO-BP 算法预测结果, 其平均绝对百分误差更是提高了 0.3%, 因此 LSTM 的预测效果优于粒子群优化的神经网络算法。

如表 5 所示, 从 2016—2020 年测试集数据对比可知, 5 种预测模型的相对平均误差分别为 16.92%、14.82%、4.06%、1.162%、0.45%。小波神经网络和 BP 神经网络的预测值不仅明显低于真实值, 而且平均相对误差接近 20%, 误差较大不太适合本次数据预测, 而优化后的 GA-BP、PSO-BP、LSTM 算法误差小于 5%, 均可以用来预测未来五年的生鲜农产品的需求量值, 明显 LSTM 物流需求预测模型的预测更有优势, 比单一的网络及优化后的网络模型的精度高。

表 4 不同模型的训练集误差数据对比
Tab.4 Comparison of training set error data of different models

算法	误差		
	RMSE	MAE	MAPE/%
小波神经网络	2 466.21	2 462.95	20.3
BP 神经网络	2 815.12	2 746.90	23.6
GA-BP 神经网络	601.47	591.68	3.9
PSO-BP 神经网络	217.43	190.42	1.3
LSTM	203.13	178.29	1.0

表 5 不同模型测试结果对比
Tab.5 Comparison of test results of different models

年份	真实值	小波神经网络		BP 神经网络		GA-BP 神经网络		PSO-BP 神经网络		LSTM	
		预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%	预测值	相对误差/%
2016	14 522.75	12 061.13	17.0	11 991.05	17.4	14 912.72	2.7	14 362.46	0.4	14 471.81	0.35
2017	14 598.87	11 948.43	18.2	12 316.94	15.6	15 204.80	4.2	14 318.12	1.9	14 591.13	0.05
2018	14 623.42	12 112.51	17.2	12 440.38	14.9	15 302.22	4.6	14 336.46	2.0	14 544.19	0.54
2019	14 452.24	12 194.22	15.6	11 604.37	19.7	15 044.81	4.1	14 453.81	0.01	14 439.99	0.09
2020	14 678.72	12 244.97	16.6	10 788.77	26.5	15 369.86	4.7	14 456.18	1.5	14 499.71	1.22
平均相对误差/%		16.92		18.82		4.06		1.162		0.45	

3.3 基于 LSTM 模型预测结果及建议

文中选用预测性能更优的 LSTM 网络模型, 将 2000—2020 年的相关影响因素的指标数据分为训练集和测试集数据, 使用训练集数据对模型进行调优, 使模型预测效果稳定后, 使用测试集数据验证模型的预测效果, 预测山东省未来 5 年的生鲜农产品冷链物流需求量, 得到的预测结果见表 6。

表 6 基于 LSTM 模型的生鲜农产品需求量的预测值
Tab.6 Predicted value of demand for fresh agricultural products based on LSTM model

年份	需求量/万 t
2021	14 580
2022	14 854
2023	15 040
2024	15 221
2025	15 470

由表 6 可知, 未来 5 年的山东省生鲜农产品的需求量将仍然呈现上升的趋势, 需求市场发展越来越快, 生鲜农产品的需求也是拉动经济发展的一大方面, 疫情期间不论政府还是企业都应加强对生鲜农产品需求市场的重视。

4 结语

分析了山东省生鲜农产品冷链物流需求预测关键影响因素, 建立了预测的指标体系。经过对 5 种模型的检测精度对比分析, 结果显示 LSTM 模型对山东省生鲜农产品冷链物流需求预测的准确率和性能更优。预测山东省未来 5 年生鲜农产品冷链仓储需求量在 14 580~15 470 万 t, 需求量将仍然呈现上升的趋势。

参考文献:

[1] 陈荣虎, 任利. 国内外物流需求预测研究趋势可视化

分析[J]. 重庆工商大学学报(社会科学版), 2021, 6(7): 1-15.

CHEN Rong-hu, REN Li. Visual Analysis of Research Trends in Logistics Demand Forecasting at Home and Abroad[J]. Journal of Chongqing Technology and Business University (Social Science Edition), 2021, 6(7): 1-15.

[2] 邓磊. 衡阳市生鲜农产品冷链物流发展策略研究[D]. 长沙: 中南林业科技大学, 2019: 8-15.

DENG Lei. The Study on Fresh Agricultural Products Cold Chain Logistics Development Strategy of Hengyang[D]. Changsha: Central South University of Forestry & Technology, 2019: 8-15.

[3] 孟庆龙, 尚静, 张艳. 苹果可溶性固形物含量的多元线性回归预测[J]. 包装工程, 2020, 41(13): 26-30.

MENG Qing-long, SHANG Jing, ZHANG Yan. Prediction for Soluble Solids Content of Apples Based on Multi Linear Regression[J]. Packaging Engineering, 2020, 41(13): 26-30.

[4] 杨玮, 张堃, 赵晶, 等. 基于 Monte Carlo 指数平滑订单预测与决策分析[J]. 包装工程, 2019, 40(5): 155-161.

YANG Wei, ZHANG Kun, ZHAO Jing, et al. Forecast and Decision Analysis of Exponential Smoothing Order Based on Monte Carlo[J]. Packaging Engineering, 2019, 40(5): 155-161. [5] HAM Y G, KIM J H, LUO Jing-jia. Deep Learning for Multi-Year ENSO Forecasts[J]. Nature, 2019, 573(7775): 568-572.

[5] KIM J, AHNI. Infectious Disease Outbreak Prediction Using Media Articles with Machine Learning Models[J]. Scientific Reports, 2021, 11: 4413.

[6] MUNKHDALAIL, PARK K H, BATBAATARE, et al. Deep Learning-Based Demand Forecasting for Korean Postal Delivery Service[J]. IEEE Access, 8: 188135-188145.

- [7] OU Shi-qi, HE Xin, JI Wei-qi, et al. Machine Learning Model to Project the Impact of COVID-19 on US Motor Gasoline Demand[J]. *Nature Energy*, 2020, 5(9): 666-673.
- [8] KRISTJANPOLLER R W, . A Stock Market Risk Forecasting Model through Integration of Switching Regime, ANFIS and GARCH Techniques[J]. *Applied Soft Computing*, 2018, 67: 106-116.
- [9] 李德奎, 杜书波, 张鹏. 基于ARIMA和LSTM的城市轨道交通延时客流预测方法比较[J]. *青岛理工大学学报*, 2021, 42(4): 135-142.
- LI De-kui, DU Shu-bo, ZHANG Peng. Comparison of Delayed Passenger Flow Forecasting Methods for Urban Rail Transit Based on ARIMA and LSTM[J]. *Journal of Qingdao University of Technology*, 2021, 42(4): 135-142.
- [10] 刘蓉晖, 赵才涛. 基于数据挖掘技术的气温敏感负荷短期预测研究[J]. *电网与清洁能源*, 2017, 33(11): 32-38.
- LIU Rong-hui, ZHAO Cai-tao. Study of the Short-Term Forecasting of the Temperature Sensitive Load Based on Data Mining Technology[J]. *Power System and Clean Energy*, 2017, 33(11): 32-38.
- [11] 马佳玉. 基于组合模型的某农场电商生鲜农产品需求量预测研究[D]. 青岛: 山东科技大学, 2020.
- MA Jia-yu. Research on Forecasting the Demand of Fresh Agricultural Products in a Farm Based on Combined Model[D]. Qingdao: Shandong University of Science and Technology, 2020.
- [12] 黄凯, 王健. 我国生鲜农产品冷链物流需求预测分析: 基于最优组合模型[J]. *武汉理工大学学报(信息与管理工程版)*, 2020, 42(6): 524-529.
- HUANG Kai, WANG Jian. Forecast and Analysis of Cold Chain Logistics Demand of Fresh Agricultural Products in China: Based on the Optimal Combination Model[J]. *Journal of Wuhan University of Technology (Information & Management Engineering)*, 2020, 42(6): 524-529.
- [13] CHEN Fu-chen, JAHANSHAHI M R. NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion[J]. *IEEE Transactions on Industrial Electronics*, 2018, 65(5): 4392-4400.
- [14] IMANIANN. Velocity Based Artificial Bee Colony Algorithm for High Dimensional Continuous Optimization Problems[J]. *Engineering Applications of Artificial Intelligence*, 2014, 36: 148-163.
- [15] 徐星, 孙光中, 田坤云. GA-BP神经网络在煤矿突水水源判别中的应用[J]. *煤炭技术*, 2018, 37(10): 172-174.
- XU Xing, SUN Guang-zhong, TIAN Kun-yun. Application of GA-BP Neural Network in Discrimination of Water Inrush Source in Coal Mine[J]. *Coal Technology*, 2018, 37(10): 172-174.
- [16] MENG An-bo, et al. Wind Speed Forecasting Based on Wavelet Packet Decomposition and Artificial Neural Networks Trained by Crisscross Optimization Algorithm[J]. *Energy Conversion and Management*, 2016, 114: 75-88.
- [17] PENG Tao, DENG Hong-wei. Comprehensive Evaluation for Sustainable Development Based on Relative Resource Carrying Capacity—A Case Study of Guiyang, Southwest China[J]. *Environmental Science and Pollution Research*, 2020, 27(16): 20090-20103.
- [18] YANG Bei-bei, YIN Kun-long, LACASSE S, et al. Time Series Analysis and Long Short-Term Memory Neural Network to Predict Landslide Displacement[J]. *Landslides*, 2019, 16(4): 677-694.
- [19] MINNS A W, HALL M J. Artificial Neural Networks as Rainfall-Runoff Models[J]. *Hydrological Sciences Journal*, 1996, 41(3): 399-417.
- [20] 李恒. 基于深度学习过拟合现象的分析[J]. *中国科技信息*, 2020(14): 90-91.
- LI Heng. Analysis of Over-Fitting Phenomenon Based on Deep Learning[J]. *China Science and Technology Information*, 2020(14): 90-91.

责任编辑: 曾钰婵