

文章编号: 1004-4353(2023)01-0077-06

基于 PCA-GRNN 模型的新能源汽车 月度销售量预测研究

谢萍萍

(黎明职业大学 智能制造工程学院, 福建 泉州 362000)

摘要: 为预测新能源汽车的月度销售量,提出了一种基于主成分分析(PCA)和广义回归神经网络(GRNN)相结合的预测模型——PCA-GRNN 模型. 首先,选取动力电池月份装车量、充电基础设施、电池级碳酸锂平均价格、交通和通信类居民消费价格指数、全国城镇调查失业率、汽车制造业工业生产者出厂价格指数等 6 个指标作为新能源汽车月度销售量的影响因子;其次,利用主成分分析方法得到可代表 6 个影响因子的 2 个主成分,并利用 Matlab 神经网络工具箱的 GRNN 神经网络函数构建了广义回归神经网络模型;最后,将 2020—2022 年间 27 个月度的统计数据分别输入到 PCA-GRNN、PCA-BP 和 PCA-Elman 模型中进行预测. 结果显示,PCA-GRNN 模型预测的新能源汽车月度销售量的平均相对误差(4.00%)低于 PCA-BP 模型和 PCA-Elman 模型预测的平均相对误差(分别为 4.77% 和 4.29%),因此 PCA-GRNN 模型在预测新能源汽车销售量方面具有一定的实用性.

关键词: 新能源汽车; 主成分分析; 广义回归神经网络; 销售量预测

中图分类号: F426.471

文献标识码: A

Research on forecast of monthly sales volume of new-energy vehicles based on PCA-GRNN model

XIE Pingping

(College of Intelligent Manufacturing Engineering, Liming Vocational University, Quanzhou 362000, China)

Abstract: To forecast the monthly sales of new-energy vehicles, a prediction model based on the principal component analysis with generalized regression neural network (PCA-GRNN) was proposed. Firstly, six indicators were selected as the influencing factors of monthly sales of new-energy vehicles, such as the monthly load of power batteries, charging infrastructure, the average price of battery-grade lithium carbonate, transportation and communication consumer price index, national urban survey unemployment rate, and industrial producer ex-factory price index of the automobile manufacturing industry. Secondly, two principal components representing most of the information of six influencing factors were obtained by PCA, and a GRNN model was constructed using the GRNN neural network function of the Matlab neural network toolbox. Finally, the statistical data of 27 months from 2020 to 2022 were input into PCA-GRNN, PCA-BP (principal component analysis-back propagation) and PCA-Elman models for forecasting, respectively. The results show that the mean relative error of the PCA-GRNN prediction model of monthly new-energy vehicle sales (4.00%) was lower than that of the PCA-BP and PCA-Elman models (4.77% and 4.29%, respectively). Therefore, the PCA-GRNN model is practicability in predicting new-energy vehicle sales.

收稿日期: 2022-05-22

基金项目: 福建省教育厅中青年教师教育科研项目(JAT201308);黎明职业大学科研团队项目(LMTD202001)

作者简介: 谢萍萍(1981—),女,硕士,讲师,研究方向为管理系统工程.

Keywords: new-energy vehicles; principal component analysis; generalized regression neural network; sales forecast

0 引言

近年来随着石油能源危机和生态环境压力的进一步加大,新能源汽车已逐步代替传统汽车,并呈快速发展趋势,如 2019 年我国新能源车的销售量为 120.6 万辆,2020 年为 136.7 万辆,2021 年为 352.1 万辆。目前,预测销售量通常采用时间序列模型^[1]、回归分析模型^[2]、灰色模型^[3]等方法,但时间序列模型和回归分析模型对解决多元非线性问题具有局限性,而灰色神经网络则存在易陷入局部极小和收敛速度慢等问题。研究显示,组合模型可以更好地挖掘数据的内部特征,提高预测的精度。例如: O.Kitapci 等^[4]利用多元回归和神经网络分析法对土耳其的汽车销售量进行了预测;周彦福等^[5]利用果蝇算法优化灰色神经网络模型后对新能源汽车的月度销售量进行了预测。广义回归神经网络(general regression neural network, GRNN)是一种建立在数理统计基础上的径向基函数网络,由于它能够较好地解决非线性问题以及提高少量数据或者数据不稳定时的预测效果,因而受到学者们的关注^[6]。目前,使用 GRNN 预测销售量的相关研究较少。梁达强^[7]和王红卫等^[8]利用 GRNN 分别对木浆和灯具的销售量进行了预测,但其研究均将所选的影响因子直接作为预测模型的输入变量(未对影响因子进行降维),因此其计算速度和预测精度存在不足。为此,本文将主成分分析(principal component analysis, PCA)方法与 GRNN 方法相结合,提出了一种 PCA-GRNN 预测模型,并对我国新能源汽车的月度销售量进行了预测。

1 相关理论介绍

1.1 主成分分析

PCA 是一种多变量统计分析方法,该方法将原来的多个变量化为少数几个主成分,以此实现降维并用以特征提取和数据压缩。PCA 的基本模型为:

$$\begin{cases} F_1 = \lambda_{11}X_1 + \lambda_{21}X_2 + \cdots + \lambda_{p1}X_p, \\ F_2 = \lambda_{12}X_1 + \lambda_{22}X_2 + \cdots + \lambda_{p2}X_p, \\ \cdots, \\ F_p = \lambda_{1p}X_1 + \lambda_{2p}X_2 + \cdots + \lambda_{pp}X_p. \end{cases} \quad (1)$$

其中: X_1, X_2, \cdots, X_p 是原始变量特征; F_1, F_2, \cdots, F_p 是新变量特征; λ 为 F 的特征值。公式(1)的特征值之间的关系为 $\lambda_{1i}^2 + \lambda_{2i}^2 + \cdots + \lambda_{pi}^2 = 1, i = 1, \cdots, p$ 。原始变量经特征向量变换后,所得的数据即为变量的主成分^[9]。

1.2 广义回归神经网络

GRNN 模型在结构上分为 4 层,如图 1 所示^[10]。第 1 层为输入层,其输入样本为 $X (X = [x_1, x_2, \cdots, x_n]^T)$ 。第 2 层为模式层,其传递函数为 $P_i (P_i = \exp[-(X - X_i)^T(X - X_i)/(2\sigma^2)], i = 1, \cdots, n, \sigma$ 为分布密度)。第 3 层为求和层,其有 2 个传递函数,其中一个函数负责算术求和($S_D = \sum_{i=1}^n P_i$),另一个函数负责加权求和($S_{Nj} = \sum_{i=1}^n y_{ij}P_i, j = 1, \cdots, k$),

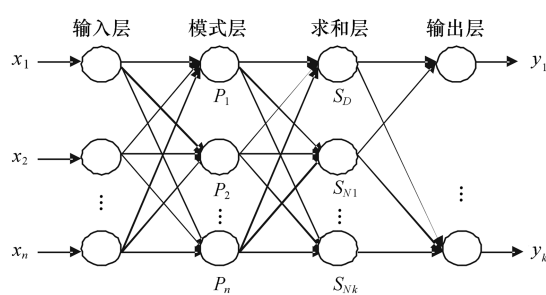


图 1 广义回归神经网络结构

y_{ij} 为第 i 个输出样本 y_i 中的第 j 个元素). 第 4 层为输出层, 其输出样本为 Y , 输出值为 S_{Nj} 的输出值除以 S_D 的输出值所得到的值, 输出层的公式为 $Y=[y_1, y_2, \dots, y_k]^T$, $y_j = \frac{S_{Nj}}{S_D}$, $j = 1, \dots, k$.

1.3 误差分析方法

本文采用绝对误差 e 和相对误差 e' 进行误差分析, 其计算公式分别为 $e = |E - E'|$, $e' = \frac{e}{E'} \times 100\%$, 其中 E 为测量值, E' 为真实值.

2 数据建模及其分析

2.1 影响因子指标的选取

影响新能源汽车月度销售量的因素较多, 本文根据文献[11-13]的研究结果选取动力电池月份装车量 $X1$ (GW · h)、充电基础设施 $X2$ (万台)、电池级碳酸锂平均价格 $X3$ (万元/t)、交通和通信类居民消费价格指数 $X4$ (上年同月 = 100)、全国城镇调查失业率 $X5$ (%)、汽车制造业工业生产者出厂价格指数 $X6$ (上年同月 = 100) 等 6 个影响因子作为构建新能源汽车月度销售量 Y (万辆) 预测模型的因子指标, 各年份及其月份的数据见表 1. 表 1 中的数据来源于国家统计局、中国汽车工业协会以及网络公开资料.

表 1 影响新能源汽车月度销售量的因子指标及其数值

年份及月份	X1	X2	X3	X4	X5	X6	Y
2020 年 1 月	2.32	53.1	5.10	100.9	5.3	99.3	4.4
2020 年 2 月	0.60	53.1	5.00	98.4	6.2	99.5	1.3
2020 年 3 月	2.77	54.2	5.12	96.2	5.9	99.5	5.3
2020 年 4 月	3.59	54.7	4.79	95.1	6.0	99.4	7.2
2020 年 5 月	3.51	55.1	4.54	94.9	5.9	99.4	8.2
2020 年 6 月	4.70	55.8	4.41	95.4	5.7	99.6	10.4
2020 年 7 月	5.02	56.6	4.48	95.6	5.7	99.6	9.8
2020 年 8 月	5.13	59.2	4.50	96.1	5.6	99.7	10.9
2020 年 9 月	6.58	60.6	4.40	96.4	5.4	99.7	13.8
2020 年 10 月	5.87	66.7	4.41	96.1	5.3	99.7	16.0
2020 年 11 月	10.61	69.5	4.48	96.1	5.2	99.6	20.0
2020 年 12 月	12.90	80.7	5.44	96.9	5.2	99.6	24.8
2021 年 1 月	8.66	81.1	7.21	95.4	5.4	99.4	19.3
2021 年 2 月	5.58	83.7	8.16	98.1	5.5	99.1	11.0
2021 年 3 月	9.00	85.1	9.00	102.7	5.3	99.2	22.6
2021 年 4 月	8.39	86.8	9.00	104.9	5.1	99.2	20.6
2021 年 5 月	9.76	88.4	9.08	105.5	5.0	99.3	21.7
2021 年 6 月	11.10	92.3	9.04	105.8	5.0	99.6	25.6
2021 年 7 月	11.29	95.0	9.16	106.9	5.1	99.5	27.1
2021 年 8 月	12.60	98.5	12.04	105.9	5.1	99.7	32.1
2021 年 9 月	15.69	104.4	17.42	105.8	4.9	99.8	35.7
2021 年 10 月	15.40	106.2	18.83	107.0	4.9	99.9	38.3
2021 年 11 月	20.80	109.2	20.31	107.6	5.0	100.2	45.0
2021 年 12 月	26.20	114.7	28.11	105.0	5.1	100.3	53.1
2022 年 1 月	16.20	117.8	38.10	105.2	5.3	100.4	43.1
2022 年 2 月	13.68	121.3	45.85	105.5	5.5	100.7	33.0
2022 年 3 月	21.40	123.2	48.50	105.8	5.8	100.7	48.4

2.2 影响因子的主成分分析

采用 SPSS 软件对因子指标进行主成分分析,其分析主要包括:因子相关系数矩阵、KMO 和 Bartlett's 球形检验、因子总方差的解释。

1) 因子相关系数矩阵. 表 2 为新能源汽车销售量影响因子的相关性矩阵. 由表中的相关系数可知,各个影响因子之间的相关性较强,因此需利用主成分分析法对其进行降维来消除各影响因子间的相关性,以减少预测模型的输入变量。

表 2 因子的相关性矩阵

相关性	X1	X2	X3	X4	X5	X6
X1	1.000	0.880	0.690	0.696	-0.713	0.672
X2	0.880	1.000	0.829	0.840	-0.697	0.631
X3	0.690	0.829	1.000	0.622	-0.298	0.850
X4	0.696	0.840	0.622	1.000	-0.719	0.402
X5	-0.713	-0.697	-0.298	-0.719	1.000	-0.207
X6	0.672	0.631	0.850	0.402	-0.207	1.000

2) KMO 和 Bartlett's 球形检验. 经检验, KMO 检验的结果为 0.685, 这表明所选因子适用于主成分分析; Bartlett's 球形检验的显著性结果为 $p = 0$ (小于 0.001), 该结果进一步表明所选因子适用于主成分分析。

3) 因子总方差的解释. 表 3 为新能源汽车销售量影响因子的总方差解释表. 由表 3 可知, 成分 1 和成分 2 的方差解释率累计为 89.625% (成分 1 的方差解释率为 71.615%, 成分 2 的方差解释率为 18.011%), 表明这两个成分包含了 6 个影响因子的大部分信息, 可以作为主成分分析的条件。

表 3 因子总方差的解释

成分	起始特征值			提取平方和载入		
	总计	变异率/%	累加/%	总计	变异率/%	累加/%
1	4.297	71.615	71.615	4.297	71.615	71.615
2	1.081	18.011	89.625	1.081	18.011	89.625
3	0.353	5.881	95.506			
4	0.140	2.328	97.834			
5	0.106	1.770	99.604			
6	0.024	0.396	100.000			

2.3 PCA-GRNN 模型的建立与预测分析

2.3.1 PCA-GRNN 模型的建立

建模时首先根据表 1 中的影响因子数据在 SPSS 软件中计算出主成分 1 和主成分 2, 然后在 Matlab 软件中调用 Matlab 神经网络工具箱中的 GRNN 神经网络函数建立 GRNN 神经网络预测模型. 按上述建立的预测模型如图 2 所示。

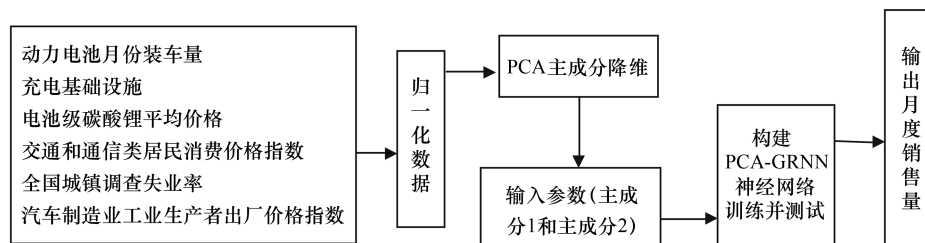


图 2 PCA-GRNN 模型的结构

2.3.2 预测方法与结果分析

预测时,将表1中的2020—2021年的24个月度数据作为模型训练数据,并将其随机分为24组;在每个组内,将前23个月度数据作为训练样本,将最后的1个月的月度数据作为测试样本.计算时,模型的分布密度值取0.1~1.0(采用试验法获得)^[14].

为检验 PCA-GRNN 模型的有效性,将 PCA-GRNN 模型与 PCA-BP、PCA-Elman 模型进行了对比. PCA-BP 模型和 PCA-Elman 模型是通过调用 Matlab 软件中的 BP 和 Elman 神经网络函数建立的,其中输入层为主成分1和主成分2,输出层为月度销售量,隐含层的节点数为2. PCA-BP 模型中的隐含层的激活函数分别为 tansig 函数和 purelin 函数,输出层函数为 trainlm 函数; PCA-Elman 模型中的隐含层的激活函数分别为 tansig 函数和 purelin 函数,输出层函数为 traingdm 函数. 3个模型均采用相同的月度数据进行学习训练. 3个模型完成训练后,以2022年1—3月的样本数据对其进行预测和性能对比.图3为3个模型的预测结果.由图可以看出,3个模型预测的销量趋势基本一致,即总体呈现稳步增长趋势.

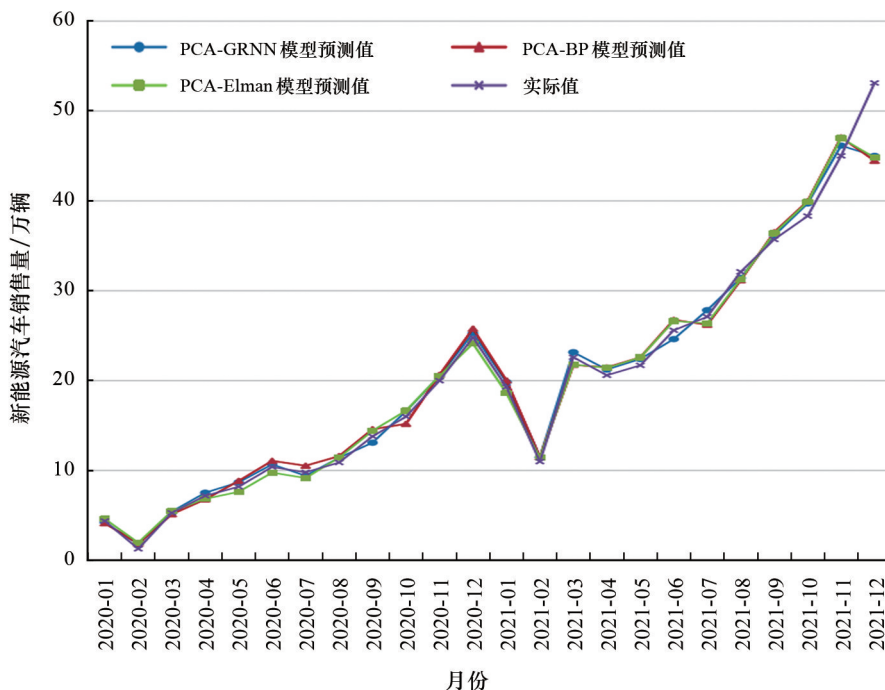


图3 3种模型预测新能源汽车月度销售量的预测值与实际值的对比

图4为3种模型预测的新能源汽车月度销售量的相对误差(预测数据的月份为2020年1月—2021年12月).由图4可见,3个模型预测的相对误差除2020年2月和2021年12月超过10%外,其余月份均在10%以内.2020年2月出现预测值相对误差较大的主要原因是销售受到了新冠疫情的影响(销售量仅达到1.3万辆);2021年12月出现预测值相对误差较大的主要原因是销售受到了新能源汽车财政补贴政策调整预期的影响(销售量达到53.1万辆).经计算,PCA-GRNN、PCA-BP和PCA-Elman模型预测的新能源汽车月度销售量的平均相对误差分别为5.34%、6.78%和6.42%.该结果表明,PCA-GRNN模型的预测效果优于PCA-BP和PCA-Elman模型.

表4为3种模型对2022年1—3月的新能源汽车月度销售量的预测结果.由表4可知,PCA-GRNN、PCA-BP和PCA-Elman模型预测的平均相对误差分别为4.00%、4.77%和4.29%,该结果进一步表明PCA-GRNN模型的预测效果优于PCA-BP和PCA-Elman模型.

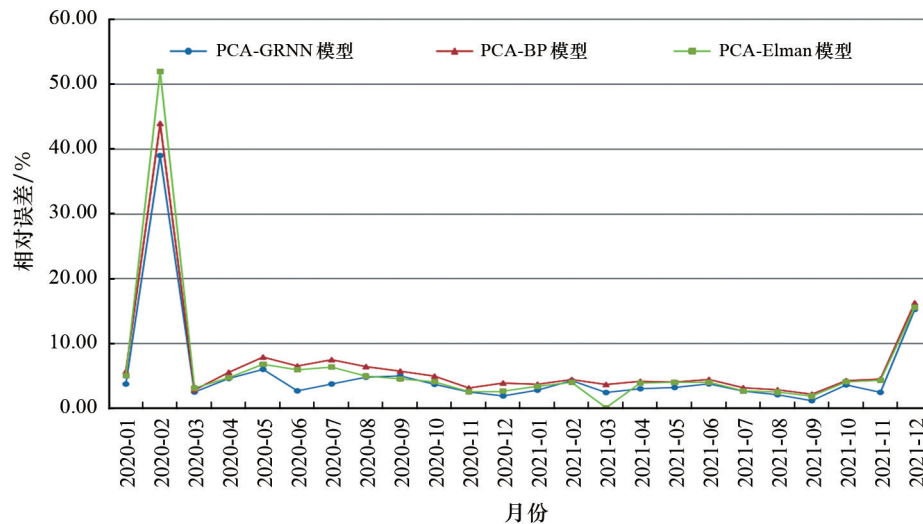


图 4 3 种模型预测新能源汽车月度销售量的相对误差

表 4 3 种模型对 2022 年 1—3 月新能源汽车销售量的预测结果

年份及月份	实际值/ 万辆	PCA-GRNN		PCA-BP		PCA-Elman	
		预测值/万辆	相对误差/%	预测值/万辆	相对误差/%	预测值/万辆	相对误差/%
2022 年 1 月	43.1	44.99	4.38	40.64	5.71	45.10	4.64
2022 年 2 月	33.0	34.52	4.60	34.68	5.10	34.59	4.81
2022 年 3 月	48.4	49.87	3.03	50.09	3.50	46.74	3.42

3 结论

利用本文提出的 PCA-GRNN 模型对 2022 年 1—3 月的新能源汽车月度销售量进行预测显示,其平均相对误差为 4.00%, 低于 PCA-BP 和 PCA-Elman 模型预测的平均相对误差(分别为 4.77% 和 4.29%), 因此本文提出的预测模型具有一定的实用价值. 由于影响销售的因素较多, 因此在今后的研究中笔者将增加其他因素(如财政补贴标准和汽车芯片产能等指标)以及采用遗传算法等来进一步提高本文模型的适用性.

参考文献:

- [1] 缪辉,唐晨添,罗露璐. 基于 ARIMA 模型的新能源汽车销量预测[J]. 企业科技与发展, 2020(10):97-98.
- [2] 张双妮. 基于多元回归模型的新能源汽车市场发展趋势预测[J]. 决策探索(中), 2019(1):77.
- [3] 余祥宽,廖秋明. 我国新能源汽车销售量预测的数学模型研究[J]. 智库时代, 2018(34):142.
- [4] KITAPCI O, ÖZEKICIOĞLU H, KAYNAR O, et al. The effect of economic policies applied in Turkey to the sale of automobiles: multiple regression and neural network analysis[J]. Procedia-Social and Behavioral Sciences, 2014 (148):653-661.
- [5] 周彦福,王红蕾. 我国新能源汽车月度销售量预测模型研究[J]. 软件导刊, 2019, 18(8):149-153.
- [6] 张山山. 基于 PCA-BA-GRNN 模型的公共自行车需求预测[D]. 兰州:兰州大学, 2018.
- [7] 梁达强. 基于神经网络的 K 公司木浆销量预测[D]. 上海:上海交通大学, 2016.
- [8] 王红卫,林健良. 基于改进的 GRNN 的销量预测[J]. 计算机工程与科学, 2010, 33(1):153-155.
- [9] 韩小孩,张耀辉,孙福军,等. 基于主成分分析的指标权重确定方法[J]. 四川兵工学报, 2012, 33(10):124-126.
- [10] 缪志刚. 基于 BP 神经网络的销售预测模型[D]. 苏州:苏州大学, 2007.
- [11] 于焱,苑鑫艺. 我国新能源汽车市场销量预测[J]. 内燃机与配件, 2021(6):179-180.
- [12] 范萌萌. 新能源汽车消费行为影响因素分析[D]. 济南:齐鲁工业大学, 2020.
- [13] 彭华. 中国新能源汽车产业发展及空间布局研究[D]. 长春:吉林大学, 2019.
- [14] 郝建浩,唐德善,尹笋,等. 基于广义回归神经网络模型的径流预测研究[J]. 水电能源科学, 2016, 34(12):49-52.