

文章编号: 1004-4353(2023)01-0053-08

区域视角下我国老龄化人口的组合预测模型

孙丽苹¹, 袁宏俊¹, 胡凌云²

(1. 安徽财经大学 统计与应用数学学院; 2. 安徽财经大学 管理科学与工程学院; 安徽 蚌埠 233030)

摘要: 为提高预测老龄人口的精度,以残差修正 GM(1,1)模型、灰色-BP 神经网络模型和 Logistic 预测模型作为单项模型,建立了一种基于相关系数的诱导有序加权平均(IOWA)算子组合预测模型. 利用该模型对 2000—2020 年我国东部、中部、西部和东北部地区及全国老龄(≥ 65 岁)人口数量进行预测显示,该组合预测模型的预测效果显著优于上述 3 个单项预测模型,表明该模型能够有效地提高老龄人口的预测精度. 利用该组合预测模型对未来 10 年我国东部、中部、西部和东北部地区及全国老龄人口数进行预测显示:我国老龄人口数总体呈现逐步增长态势,同时区域间的老龄化进程差异也不断增大,其中东部地区老龄人口增长得相对较快,中部地区则增长得相对较慢. 该预测结果可为我国老龄人口政策的制定提供参考.

关键词: 老龄人口; 单项预测模型; IOWA 算子; 组合预测模型

中图分类号: F224

文献标识码: A

Combinational forecasting model of aging population from a regional perspective

SUN Liping¹, YUAN Hongjun¹, HU Lingyun²

(1. School of Statistics and Applied Mathematics, Anhui University of Finance and Economics; 2. School of Management Science and Engineering, Anhui University of Finance and Economics; Bengbu 233030, China)

Abstract: In order to improve the predicting accuracy of the aging population, a combinational forecasting model which is based on induced ordered weighted averaging (IOWA) operators and correlation coefficient is established by using the residual modified GM(1,1) model, the gray-BP neural network model and the Logistic forecasting model as the single model. Using this model forecast the aging population (≥ 65 years old) of the eastern, central, western and northeastern regions and the whole country for 2000–2020 in China, the forecasting result shows that the predicting effect of the combinational forecasting model is significantly better than that of the above three single forecasting models, indicating the model can effectively improve the predicting accuracy of the aging population. Finally, the combinational forecasting model is used to predict the number of ageing population in the eastern, central, western and northeastern parts of China and the whole country in the next 10 years; the number of ageing population in China generally shows a gradual growth trend, and the differences in the aging process between regions are also increasing, of which the ageing population in the east grows relatively fast and the growth in the central region is relatively slow. The forecasting result can provide a reference for the formulation of Chinese aging population policy.

Keywords: ageing population; single-term forecasting model; IOWA operator; combinational forecasting model

收稿日期: 2022-07-12

基金项目: 安徽省哲学社会科学规划项目(AHSKY2020D42); 安徽财经大学重大科研基金(ACKYA21004); 安徽财经大学研究生科研创新基金(ACYC2021400)

第一作者: 孙丽苹(1996—),女,硕士研究生,研究方向为经济预测理论与方法.

通信作者: 袁宏俊(1978—),男,教授,研究方向为预测理论与方法.

统计显示,2020 年我国人口(14.11 亿人)中 65 岁及以上人口总数为 1.90 亿人(占比为 13.5%)^[1],这表明我国已经进入老龄化社会.人口老龄化不仅会加大社会保障和公共服务压力,而且会持续影响社会活力、创新动力和经济潜在增长率,因此研究老龄人口变化具有重要意义.传统预测人口的方法主要有 GM(1,1)模型^[2]、Logistic 模型^[3]、BP 神经网络模型^[4]、队列要素法^[5]等,这些单一模型虽然具有自我学习能力强和使用数据少等优点,但由于它们难以同时捕获数据序列的线性及非线性特征,因此在预测波动性强的数据序列时存在精度较低的问题.为此,一些学者提出了基于不同准则的定权组合模型和变权组合模型,如徐翔燕等构建的 GM(1,1)-SVM 组合预测模型^[6]、龙会典等构建的 GM(1,1)-Markov 链组合预测模型^[7]、袁宏俊等构建的 COWG-WPA 算子区间型组合预测模型等^[8-10],研究表明这些组合模型的预测效果均优于单项预测模型.基于上述研究,本文以灰色-BP 神经网络模型、残差修正 GM(1,1)模型和 Logistic 模型作为单项预测模型,构建了一种基于相关系数的诱导有序加权平均(IOWA)算子组合预测模型,并利用该模型预测了 2021—2030 年我国东部、中部、西部和东北部地区及全国的老龄人口发展趋势.

1 老龄人口总量的单项预测模型

1.1 残差修正 GM(1,1)模型

灰色理论^[11]由邓聚龙教授首次提出,其中 GM 系列模型因具有使用数据少和计算简单等优点被广泛应用于经济、农业及交通运输等多个领域.由于修正的灰色预测模型优于 GM(1,1)预测模型^[12],因此本文选用修正的灰色预测模型对老龄人进行预测.构建修正的灰色预测模型的步骤如下:

步骤 1 假设老龄人口的原始非负序列 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 则由此建立的基于原始序列 $x^{(0)}$ 的 GM(1,1) 模型为 $\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b_1}{a_1} \right] e^{-a_1 k} + \frac{b_1}{a_1}$, 其中 $k \in [1, n]$.

步骤 2 计算序列 $x^{(0)}$ 的预测值序列 $\hat{x}^{(0)}$. 利用步骤 1 中的 GM(1,1) 模型对序列 $\hat{x}^{(1)}$ 进行一次累减, 由此可得到序列 $x^{(0)}$ 的预测值序列 $\hat{x}^{(0)}$:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad (1)$$

$$\hat{x}^{(0)}(k+1) = (1 - e^{a_1}) \left[x^{(0)}(1) - \frac{b_1}{a_1} \right] e^{-a_1 k}. \quad (2)$$

步骤 3 计算残差序列 $E^{(0)}$ 的预测值序列 $\hat{E}^{(0)}$. 利用式(2)对序列 $x^{(0)}$ 作差得残差序列 $E^{(0)}$ 的预测值, 并选取其尾部的数值作为残差子序列 $\epsilon^{(0)}$ ($\epsilon^{(0)} = \{E^{(0)}(k_0), E^{(0)}(k_0+1), \dots, E^{(0)}(n)\}$). 当序列 $\epsilon^{(0)}$ 中的数全为正或全为负时, 则直接建模; 当序列 $\epsilon^{(0)}$ 中的数为正负数, 则在序列中加上最小负数的绝对值后建模(在求得结果后减去该绝对值). 当序列 $\epsilon^{(0)}$ 中的数全为正或全为负时, 建立的基于序列 $\epsilon^{(0)}$ 的 GM(1,1) 模型为:

$$E^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), \quad (3)$$

$$\hat{E}^{(1)}(k+1) = \left[E^{(0)}(k_0) - \frac{b_2}{a_2} \right] e^{-a_2 k} + \frac{b_2}{a_2}, \quad (4)$$

$$\hat{E}^{(0)}(k+1) = \hat{E}^{(1)}(k+1) - \hat{E}^{(1)}(k), \quad (5)$$

其中 $k \in [k_0, n]$.

步骤 4 计算序列 $x^{(0)}$ 的最终预测值. 利用式(6)计算 $\hat{x}^{(0)}$ 的残差修正 GM(1,1) 模型.

$$\hat{x}^{(0)}(k+1) = \begin{cases} (1 - e^{a_1}) \left[x^{(0)}(1) - \frac{b_1}{a_1} \right] e^{-a_1 k}, & k \leq k_0; \\ (1 - e^{a_1}) \left[x^{(0)}(1) - \frac{b_1}{a_1} \right] e^{-a_1 k} + (1 - e^{a_2}) \left[E^{(0)}(k_0) - \frac{b_2}{a_2} \right] e^{-a_2 k}, & k > k_0, \end{cases} \quad (6)$$

其中 k_0 指未参加残差修正的个数.

1.2 灰色-BP 神经网络模型

BP 神经网络模型^[13]是一种能够模仿人脑结构及其功能的信息处理系统,其运算方式主要为信息前向传播和误差反向传播.由于该模型具有超强的自学习与自组织能力以及具有良好的非线性映射、容错和泛化等能力,因此被广泛应用于经济、环境科学及自动化技术等领域.研究显示,将灰色预测模型和 BP 神经网络模型相结合可有效提高预测的准确性^[14],因此本文利用 GM(1,1)和 BP 神经网络模型构建灰色-BP 神经网络模型,构建步骤为:

步骤 1 建立 GM(1,1)模型并获取预测值 $\hat{x}^{(0)}$.

步骤 2 构建 BP 神经网络模型的输入矩阵 \mathbf{P} 和输出向量 \mathbf{y} .将灰色预测值的前 3 年数据作为第 1 组样本的输入向量,再用第 4 年数据替换第 1 年数据后的数据作为第 2 组样本的输入向量,如此迭代即可得输入矩阵 \mathbf{P} .将实际值的第 4 年数据作为第 1 组样本的输出值,将实际值的第 5 年数据作为第 2 组样本的输出值,以此类推即可得到输出向量 \mathbf{y} .

步骤 3 利用输入矩阵和输出向量构建 BP 神经网络模型.

步骤 4 利用灰色-BP 神经网络模型计算预测值,并对该模型的外推性进行验证.若误差函数值满足模型的外推性条件,则对未来数据进行预测;否则返回步骤 3,调整部分参数后重新构建 BP 神经网络模型.

1.3 Logistic 模型

Logistic 模型是基于 Malthus 模型改进的一种模型.该模型的两个假定条件为:一是假定人口增长率与人口数量之间为减函数关系,即 $r(x) = r - s \cdot x$;二是假定在有限的自然资源条件下所能容纳的最

大人口数为 x_m ,即 $s = \frac{r}{x_m}$.Logistic 模型的基本方程为 $\begin{cases} \frac{dx}{dt} = r(x) \cdot x, \\ x(t_0) = x_0. \end{cases}$ 通过计算得 Logistic 模型为

$x_t = \frac{x_m}{1 + (x_m/x_0 - 1)e^{-r(t-t_0)}}$,其中 r 为人口固有增长率, $r(x)$ 为人口增长率, x 为人口数量, x_0 为 t_0 时刻的人口总数.

2 基于相关系数和 IOWA 算子的老龄人口的组合预测模型

2.1 模型的建立

本文构建基于相关系数和 IOWA 算子的组合预测模型对不同区域及全国老龄人口总量进行预测.构建该模型的基本思想是依据不同单项预测模型在不同时刻上的预测精度对其赋予不同的权重,从而提高预测的准确性.设有 n 个二维数组 $\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \dots, \langle v_n, a_n \rangle$,且 $f_L(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \dots,$

$\langle v_n, a_n \rangle) = \sum_{i=1}^n l_i a_{v\text{-index}(i)}$,则称 f_L 是由 v_1, v_2, \dots, v_n 所产生的 n 维 IOWA 算子.其中: v_i 为诱导变量;

$v\text{-index}(i)$ 是 v_i 从大到小排列的第 i 个数的下标; $\mathbf{L} = (l_1, l_2, \dots, l_n)^T$ 为加权向量,满足 $\sum_{i=1}^n l_i = 1, 0 \leq$

$l_i \leq 1$ ^[15].本文提出的组合预测模型的步骤如下:

步骤 1 计算第 i 种单项预测方法的预测精度 a_{it} ,其计算公式为:

$$a_{it} = \begin{cases} 1 - |(x_t - x_{it})/x_t|, & |(x_t - x_{it})/x_t| < 1; \\ 0, & |(x_t - x_{it})/x_t| \geq 1. \end{cases}$$

其中: $i = 1, 2, \dots, n$; $t = 1, 2, \dots, N$; x_t 为老龄人口实际值; x_{it} 为第 i 种单项预测方法的预测值.

步骤 2 基于 IOWA 算子计算组合预测值 \hat{x}_t ,其计算公式为 $\hat{x}_t = \sum_{i=1}^n l_i x_{a\text{-index}(it)}$.

步骤 3 构建以相关系数最大化为准则的最优组合预测模型,权重向量 \mathbf{L} 的计算公式为:

$$\max R(l_1, l_2, \dots, l_n) = \frac{\sum_{t=1}^N (x_t - \bar{x})(\hat{x}_t - \bar{\hat{x}}_t)}{\sqrt{\sum_{t=1}^N (x_t - \bar{x})^2} \cdot \sqrt{\sum_{t=1}^N (\hat{x}_t - \bar{\hat{x}}_t)^2}}; \text{ s. t. } \begin{cases} l_1 + l_2 + \dots + l_n = 1, \\ l_i \geq 0. \end{cases}$$

其中 $\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$, $\bar{\hat{x}}_t = \frac{1}{N} \sum_{t=1}^N \hat{x}_t$.

步骤 4 计算未来 k 期第 i 种单项预测方法的预测精度 a_{ik} , 其计算公式为 $a_{ik} = \frac{1}{k} \sum_{t=N-k+1}^N a_{it}$, 其中 a_{it} 为第 i 种单项预测方法在 t 时刻的预测精度.

步骤 5 计算未来 k 期的组合预测值,其计算公式为:

$$f_L(\langle a_{1,N+1}, x_{1,N+1} \rangle, \langle a_{2,N+2}, x_{2,N+2} \rangle, \dots, \langle a_{n,N+k}, x_{n,N+k} \rangle) = \sum_{i=N+1}^{N+k} l_i x_{a-\text{index}(it)}.$$

2.2 预测误差评价指标体系

为了准确地描述不同模型的预测效果,本文选取平方和误差(SSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、均方误差(MSE)、均方百分比误差(MSPE)对预测模型进行评价,其误差计算公式分别为:

$$SSE = \sum_{t=1}^N (x_t - \hat{x}_t)^2, MSE = \frac{1}{N} \sqrt{\sum_{t=1}^N (x_t - \hat{x}_t)^2}, MAE = \frac{1}{N} \sum_{t=1}^N |x_t - \hat{x}_t|, MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{x_t - \hat{x}_t}{x_t} \right|, MSPE = \frac{1}{N} \sqrt{\sum_{t=1}^N \left(\frac{x_t - \hat{x}_t}{x_t} \right)^2},$$

其中 x_t 和 \hat{x}_t 分别为 t 时刻的实际值和预测值.

3 老龄人口预测的实证研究

3.1 数据来源

本文选取 2000—2020 年《中国人口和就业统计年鉴》中 31 个省、自治区、直辖市的 65 岁及以上人口数作为研究数据(见表 1),运用灰色-BP 神经网络模型、残差修正 GM(1,1)模型和 Logistic 模型进行组合预测.

3.2 老龄人口数的单项预测

3.2.1 残差修正 GM(1,1)模型的预测结果

以预测东部地区老龄人口为例.首先,根据该地区老龄人口的实际值计算并构建老龄人口序列的 GM(1,1)模型,构建的模型为 $\hat{x}^{(0)}(k+1) = 5125.8053e^{0.0199k} - 1953$, 记作模型 1. 其次,利用残差序列对模型 1 进行修正,并以 $k_0 = 9$ 作为临界值,将残差序列分为两部分. 再次,对残差序列后半部分进行 GM(1,1)建模,得修正模型为 $\hat{E}^{(0)}(k+1) = 3024.8198e^{0.0307k} - 3778.4$, 记作模型 2. 最后,根据模型 1 和模型 2 得修正的 $\hat{x}^{(0)}(k+1)$ 的时间响应式为:

$$\hat{x}^{(0)}(k+1) = \begin{cases} 5125.8053e^{0.0199k} - 1953, & k \leq 9; \\ 5125.8053e^{0.0199k} + 3024.8198e^{0.0307k} - 5731.4, & k > 9. \end{cases}$$

运用所得的修正模型对各地区及全国老龄人口数进行预测的结果见表 2.

3.2.2 灰色-BP 神经网络模型的预测结果

以 2003—2018 年数据为训练集,以 2019—2020 年数据为测试集构建灰色-BP 神经网络模型. 其中:输入层神经元数为 3,输出层为 1,隐含层的取值范围为 3~12;输入层和隐含层之间的激活函数为 tansig 函数,隐含层与输出层之间的激活函数为 logsig 函数,训练函数为 trainlm 函数. 利用 Matlab 软

件计算得到的各区域及全国老龄人口数的预测结果见表 3.

表 1 各地区及全国 65 岁及以上老龄人口的实际数量					万人
年份	东部	中部	西部	东北部	全国
2000	2 976.6	2 409.5	2 277.5	693.2	8 821.0
2001	3 440.4	2 634.0	2 554.2	760.1	9 062.0
2002	3 537.7	2 823.9	2 753.5	763.4	9 377.0
2003	3 861.1	2 861.0	2 833.5	852.6	9 692.0
2004	3 776.5	2 954.8	2 909.7	860.3	9 857.0
2005	3 828.6	3 177.0	3 198.3	911.1	10 055.0
2006	3 900.0	3 211.9	3 238.8	985.5	10 419.0
2007	4 058.1	3 209.5	3 288.4	1 039.5	10 636.0
2008	4 177.9	3 283.5	3 385.9	1 087.4	10 956.0
2009	4 282.1	3 398.2	3 502.5	1 074.1	11 307.0
2010	3 891.1	3 165.8	3 245.5	998.4	11 894.0
2011	3 888.7	3 345.9	3 392.6	1 017.8	12 277.0
2012	4 074.2	3 458.7	3 485.0	987.7	12 777.0
2013	4 306.7	3 453.5	3 627.6	1 057.0	13 262.0
2014	4 454.4	3 519.9	3 876.8	1 173.9	13 902.0
2015	4 706.7	3 790.1	3 817.5	1 277.5	14 524.0
2016	4 943.1	3 924.2	3 960.5	1 329.0	15 037.0
2017	5 224.5	4 189.9	4 104.5	1 406.3	15 961.0
2018	5 582.7	4 333.8	4 348.1	1 448.2	16 724.0
2019	5 936.4	4 575.2	4 574.8	1 567.5	17 767.0
2020	6 313.9	5 070.3	5 025.8	1 614.7	19 064.0

表 2 利用残差修正 GM(1,1)模型预测各地区及全国老龄人口的结果						万人
年份	东部	中部	西部	东北部	全国	
2003	3 488.1	2 785.6	2 811.2	815.4	9 182.6	
2004	3 597.5	2 870.4	2 897.2	848.0	9 554.9	
2005	3 709.0	2 957.2	2 985.4	881.7	9 942.3	
2006	3 822.8	3 046.2	3 075.8	916.4	10 345.3	
2007	3 938.9	3 137.4	3 168.5	952.3	10 764.8	
2008	4 057.3	3 230.9	3 263.5	989.4	11 201.2	
2009	4 282.1	3 326.6	3 502.5	1 074.1	11 307.0	
2010	3 642.4	3 424.8	3 186.5	876.6	11 532.5	
2011	3 865.5	3 525.4	3 329.0	953.1	12 116.8	
2012	4 094.2	3 458.7	3 474.6	1 031.4	12 724.2	
2013	4 328.6	3 358.5	3 623.7	1 111.6	13 355.9	
2014	4 568.8	3 552.4	3 776.1	1 193.7	14 012.7	
2015	4 815.1	3 755.3	3 932.0	1 277.7	14 695.7	
2016	5 067.6	3 967.6	4 091.5	1 363.8	15 405.8	
2017	5 326.4	4 189.9	4 254.7	1 451.9	16 144.3	
2018	5 591.7	4 422.9	4 421.6	1 542.1	16 912.1	
2019	5 863.7	4 667.1	4 592.4	1 634.5	17 710.5	
2020	6 142.6	4 923.3	4 767.0	1 729.1	18 540.7	

表 3 利用灰色-BP 神经网络模型预测各地区及全国老龄人口的结果						万人
年份	东部	中部	西部	东北部	全国	
2003	3 806.1	2 886.8	2 830.7	861.7	10 131.7	
2004	3 843.0	2 959.4	2 933.7	853.1	10 021.1	
2005	3 878.6	3 147.6	3 226.2	902.4	10 169.7	
2006	3 964.7	3 249.7	3 249.1	981.6	10 341.6	
2007	4 104.1	3 281.7	3 331.6	1 047.0	10 547.7	
2008	4 207.5	3 297.1	3 473.6	1 077.6	10 806.5	
2009	4 205.4	3 316.8	3 489.1	1 075.1	11 148.1	
2010	4 085.3	3 319.4	3 392.0	1 030.5	11 615.3	
2011	3 914.5	3 334.2	3 401.7	971.4	12 242.5	
2012	4 019.1	3 377.9	3 574.3	995.5	12 989.3	
2013	4 336.1	3 466.6	3 731.8	1 082.0	13 684.2	
2014	4 519.7	3 594.8	3 839.8	1 168.0	14 198.7	
2015	4 697.1	3 753.0	3 926.5	1 251.3	14 673.7	
2016	4 952.0	3 944.4	4 016.9	1 340.3	15 232.5	
2017	5 247.7	4 158.2	4 158.7	1 401.2	15 870.2	
2018	5 584.3	4 392.7	4 389.8	1 454.9	16 583.0	
2019	5 906.2	4 637.7	4 722.8	1 523.6	17 383.6	
2020	6 201.7	4 894.9	5 207.9	1 611.7	18 285.5	

3.2.3 Logistic 模型的构建及其预测

本文采用刘磊提出的等距三点式分组法^[16]计算 Logistic 模型的参数并以此构建模型. 以预测东部地区老龄人口为例: 将老龄人口实际值代入文献^[16]中构建的代数求解公式进行计算得 $a_3 = -0.012$, $b_3 = -0.0896$, $k = 8.0461$. 由该计算结果得到的 Logistic 模型为 $x_t = \frac{8.046}{1 - 0.012e^{0.0896t}}$. 利用上述方法所构建的模型预测各地区及全国老龄人口的结果见表 4.

3.3 基于相关系数和 IOWA 算子组合模型的预测结果

以预测西部地区老龄人口为例建立的基于相关系数和 IOWA 算子的组合预测模型为:

$$R = \frac{2495.8l_1 + 2415.3l_2 + 2238.3l_3}{\sqrt{\mathbf{L}^T \mathbf{C} \mathbf{L}}}, \quad (7)$$

其中权重向量 $\mathbf{L} = (l_1, l_2, l_3)^T$. 在 t 时刻, l_1 为预测精度最高方法的权重, 其次依次为 l_2 、 l_3 . 预测误差

信息矩阵为 $\mathbf{C} = \begin{pmatrix} 52017 & -62862 & -151537 \\ -62862 & 231574 & 326559 \\ -151537 & 326559 & 869093 \end{pmatrix}$. 将该矩阵代入式(7)得该组合模型的最优权重向

量为: $\mathbf{L} = (0.7651, 0.1357, 0.0992)^T$. 运用该组合模型预测 t 时刻各地区及全国老龄人口数的结果见表 5.

表 4 利用 Logistic 模型预测各地区及全国
老龄人口的结果

年份	东部	中部	西部	东北部	万人
2003	3548.0	2796.8	2725.0	801.5	9340.9
2004	3591.5	2846.9	2800.5	824.4	9575.4
2005	3639.9	2900.7	2878.8	848.7	9826.3
2006	3693.8	2958.4	2960.0	874.6	10095.3
2007	3753.9	3020.3	3044.3	902.4	10384.2
2008	3821.0	3087.1	3131.8	932.1	10695.5
2009	3896.1	3158.9	3222.7	963.8	11031.8
2010	3980.3	3236.5	3317.0	998.0	11396.0
2011	4075.1	3320.4	3415.1	1034.6	11792.0
2012	4181.9	3411.2	3517.0	1074.1	12223.7
2013	4302.8	3509.7	3623.0	1116.7	12696.4
2014	4440.0	3616.7	3733.2	1162.7	13216.1
2015	4596.3	3733.3	3847.8	1212.5	13789.9
2016	4775.1	3860.5	3967.1	1266.5	14426.9
2017	4980.6	3999.5	4091.3	1325.3	15137.8
2018	5218.1	4152.0	4220.6	1389.2	15936.3
2019	5494.0	4319.6	4355.4	1459.1	16839.4
2020	5816.6	4504.3	4495.8	1535.5	17869.0

表 5 利用组合预测模型预测各地区及全国
老龄人口的结果

年份	东部	中部	西部	东北部	万人
2003	3789.0	2886.1	2817.5	857.6	9458.2
2004	3827.7	2958.7	2892.6	852.7	9954.9
2005	3866.9	3146.3	3159.0	900.6	9976.0
2006	3953.7	3248.3	3196.9	975.9	10344.8
2007	4090.7	3138.4	3281.0	1038.7	10580.0
2008	4194.2	3296.6	3411.2	1069.7	10865.1
2009	4272.1	3326.6	3472.9	1074.2	11283.4
2010	3979.3	3237.1	3224.6	1000.8	11603.0
2011	3871.3	3334.1	3396.3	1029.0	12223.8
2012	4092.3	3458.3	3490.3	998.7	12763.6
2013	4304.5	3466.9	3634.3	1084.6	13404.6
2014	4445.8	3552.7	3820.5	1167.5	14040.3
2015	4700.7	3755.3	3866.8	1275.4	14677.0
2016	4954.2	3944.6	3986.2	1342.4	15258.3
2017	5246.7	4189.7	4116.6	1405.7	15910.8
2018	5578.4	4392.9	4377.4	1449.1	16631.8
2019	5897.2	4637.9	4586.6	1533.5	17662.0
2020	6192.5	4923.1	5077.5	1605.0	18502.8

为了对各预测模型的有效性进行比较评价, 利用 SSE、MAE、MSE、MAPE、MSPE 对各预测模型的预测结果进行了检验, 结果见表 6. 由表 6 可以看出: 组合预测模型的各项评价指标值均显著优于各单项预测模型; 在 3 个单项预测模型中, 预测精度高低的顺序依次是灰色-BP 神经网络模型、残差修正 GM(1,1)模型、Logistic 模型. 这表明, 本文提出的组合模型可以有效地提高老龄人口预测的准确性.

利用本文提出的组合模型对我国东部、中部、西部和东北部地区及全国的 2021—2030 年老龄人口进行了预测, 其结果见表 7. 由表 7 可以看出, 未来 10 年内我国老龄人口数将保持增长的态势, 其中: 东

部地区的老龄人口数呈持续高增长的态势,2030 年将达到 9 860.6 万人;中部地区在 2021—2026 年呈缓慢增长趋势,在 2027—2030 年则呈先下降后不断增长的趋势,其中在 2029 年达到峰值(7 893.2 万人);西部地区在 2021—2023 年呈缓慢增长趋势,在 2023—2030 年则出现先大幅增长后缓慢增长的趋势,其中在 2030 年达到峰值(8 039.1 万人);东北部地区在 2021—2030 年始终保持平稳增长,在 2030 年达到峰值(2 114.4 万人). 这表明,我国区域间的老龄化进程差异不断增大,其中东部地区的老龄人口在未来 10 年的增长速度达到 51.9%,而中部地区仅为 24.4%.

表 6 各模型预测的评价结果

区域	评价指标	灰色-BP 神经网络模型	Logistic 模型	残差修正 GM(1,1)模型	相关系数和 IOWA 算子组合预测模型
东部	SSE	83 775.57	1 309 678.07	368 897.61	39 023.42
	MAE	52.44	231.57	111.37	34.04
	MSE	16.08	63.58	33.74	10.97
	MAPE/%	1.23	5.04	2.60	0.77
	MSPE/%	0.39	1.35	0.82	0.24
中部	SSE	91 947.00	770 649.82	252 140.49	50 482.20
	MSE	16.85	48.77	27.90	12.48
	MAE	54.70	163.21	95.22	39.23
	MAPE/%	1.48	4.38	2.76	1.05
	MSPE/%	0.45	1.23	0.83	0.31
西部	SSE	127 798.81	779 704.29	245 180.42	16 002.40
	MSE	19.86	49.06	27.51	7.03
	MAE	65.94	155.30	90.90	24.95
	MAPE/%	1.72	4.24	2.45	0.68
	MSPE/%	0.51	1.29	0.73	0.19
东北部	SSE	7 106.32	122 819.25	78 366.26	3 092.64
	MSE	4.68	19.47	15.55	3.09
	MAE	14.47	71.82	55.2	9.48
	MAPE/%	1.28	6.36	4.83	0.83
	MSPE/%	0.41	1.75	1.36	0.26
全国	SSE	1 526 847.97	6 644 002.00	1 137 352.17	624 426.09
	MSE	68.65	143.20	59.25	43.90
	MAE	232.00	546.56	202.37	137.60
	MAPE/%	1.74	3.95	1.60	1.02
	MSPE/%	0.49	0.97	0.48	0.30

表 7 2021—2030 年不同地区和全国老龄人口的组合预测结果 万人

年份	东部	中部	西部	东北部	全国
2021	6 491.5	5 191.8	5 592.6	1 694.9	19 388.2
2022	6 831.9	5 385.1	5 285.4	1 795.7	20 319.5
2023	7 212.9	5 612.3	5 535.7	1 839.3	21 297.3
2024	7 635.6	5 807.4	7 189.4	1 901.8	22 320.9
2025	8 063.2	5 977.6	7 529.1	1 934.6	23 388.0
2026	8 487.3	6 121.4	7 753.7	1 964.0	24 495.3
2027	8 882.6	7 112.1	7 889.4	1 996.5	25 638.4
2028	9 239.5	6 329.5	7 970.0	2 030.9	26 812.5
2029	9 559.7	7 893.2	8 013.3	2 066.2	28 012.9
2030	9 860.6	6 457.3	8 039.1	2 114.4	29 236.3

4 结论

研究表明,本文提出的基于相关系数的 IOWA 算子组合预测模型的预测效果显著优于残差修正 GM(1,1)模型、灰色-BP 神经网络模型和 Logistic 模型,因此可利用本文模型对老龄人口进行预测.对我国东部、中部、西部和东北部地区及全国未来 10 年老龄人口数进行预测显示:我国老龄人口整体呈持续增长趋势,同时区域间的老龄化进程差异不断增大,其中东部地区老龄人口增长速度相对较快,中部地区则相对较慢.该研究结果可为各区域的老龄人口政策的制定提供参考.由于老龄人口数会受到经济发展、人口政策及医疗技术等各种因素的影响,因此为了达到更好的预测效果,在今后研究中我们拟将更多的因素加入到预测模型中,如人口出生率、人均地区生产总值等.

参考文献:

- [1] 中华人民共和国国家统计局. 中国统计年鉴[M]. 北京:中国统计出版社,2021.
- [2] 张振华. 基于灰色 GM(1,1)模型的城市人口老龄化预测[J]. 统计与决策,2015,439(19):76-79.
- [3] 王学保,蔡果兰. Logistic 模型的参数估计及人口预测[J]. 北京工商大学学报(自然科学版),2009,27(6):75-78.
- [4] 史佳,冀巨海. 基于 BP 神经网络的黄河中下游人口预测研究[J]. 科技管理研究,2014,34(6):245-250.
- [5] 李爱华,王迪文. 队列要素法在人口预测中的应用[J]. 统计与决策,2021,37(22):36-40.
- [6] 徐翔燕,侯瑞环. 基于 GM(1,1)-SVM 组合模型的中长期人口预测研究[J]. 计算机科学,2020,47(S1):485-487.
- [7] 龙会典,严广乐. 基于改进的 GM(1,1)-Markov 链组合模型广东省单位 GDP 能耗预测[J]. 数理统计与管理,2017,36(2):200-207.
- [8] 袁宏俊,胡凌云. 基于连续区间有序几何加权平均 Power 算子的区间型组合预测模型[J]. 数理统计与管理,2019,38(2):270-280.
- [9] 袁宏俊,陈华友,胡凌云. 基于指数支撑度的最优组合预测模型及其性质研究[J]. 应用概率统计,2012,28(2):150-160.
- [10] 袁宏俊,胡凌云,张敏. 基于广义诱导连续区间有序函数比例加权平均算子的区间型组合预测模型[J]. 延边大学学报(自然科学版),2018,44(4):315-321.
- [11] 刘思峰,杨英杰,吴利丰. 灰色系统理论及其应用[M]. 北京:科学出版社,1999:113-116.
- [12] 张新生,赵梦旭,王小完. 尾段残差修正 GM(1,1)模型在管道腐蚀预测中的应用[J]. 中国安全科学学报,2017,27(1):65-70.
- [13] 陈强. 机器学习及 R 应用[M]. 北京:高等教育出版社,2020.
- [14] 王守相,张娜. 基于灰色神经网络组合模型的光伏短期出力预测[J]. 电力系统自动化,2012,36(19):37-41.
- [15] 陈华友,程蕾,张倩. 基于相关系数的 IOWA 算子最优组合预测模型[C]//第四届中国不确定系统年会论文集. 桂林:中国运筹学会,2006:242-248.
- [16] 刘磊. 逻辑增长曲线模型的一种参数估计方法[J]. 湖北工业大学学报,2008,23(1):83-84.