

## 基于广义回归神经网络的人口预测模型

夏福涛

贵州医科大学 贵州贵阳

**【摘要】**人口问题一直是社会发展问题的核心，是制约社会向前发展的关键因素，人口增长的过快或过慢，都会对世界发展带来一定的损失，正确的引导是解决人口问题的重要手段，为建立人口预测的数学模型，分别预测 2100 年底中国、印度和全球的总人口数，并分析人口的变化趋势。通过建立广义回归神经网络模型，使用 Matlab 进行编程预测，分别对 1950~2021 年的中国、印度和全球的总人口数进行预测，并与实际人口数量进行比较，证明模型的可行性。再进一步预测 2022~2100 年中国、印度和全球的人口总数，对结果进行可视化处理。

**【关键词】**广义回归神经网络; Matlab; 人口预测

**【收稿日期】**2023 年 4 月 6 日 **【出刊日期】**2023 年 5 月 21 日 **【DOI】**10.12208/j.aics.20230022

### A population prediction model based on generalized regression neural network

Futao Xia

Guizhou Medical University, Guiyang, Guizhou

**【Abstract】** Population problem has always been the core of social development and a key factor restricting social development. Too fast or too slow population growth will bring certain losses to world development. Correct guidance is an important means to solve the population problem. In order to establish the mathematical model of population prediction, the total population of China, India and the world at the end of 2100 is predicted respectively, and the changing trend of population is analyzed. By establishing a generalized regression neural network model and using Matlab for programming prediction, the total population of China, India and the whole world from 1950 to 2021 are predicted respectively, and the actual population is compared to prove the feasibility of the model. Furthermore, the total population of China, India and the world in 2022-2100 is predicted, and the results are visualized.

**【Keywords】** Generalized regression neural network; Matlab; Population forecast

### 1 引言

当前世界人口数量是二十世纪中期的三倍多。1950 年，全球人口约 25 亿，到 2022 年 11 月中旬，这一数字已达 80 亿，自 2010 年以来增加 10 亿，自 1998 年以来增加 20 亿。预计在未来 30 年，世界人口将增加近 20 亿，从目前的 80 亿增至 2050 年的 97 亿，并可能在 2080 年代中期达到近 104 亿的峰值。通过绘制全球 1950~2021 年总人口随时间变化的趋势折线图，分析整体人口的变化趋势<sup>[1]</sup>。

人口预测一直以来都是社会规划、政策制定和经济发展中的重要课题。对于政府、学者和企业而言，准确预测未来人口的规模、结构和分布对于制

定合理的发展战略和决策至关重要。然而，人口的预测是一个复杂而具有挑战性的任务，受到多种因素的影响，包括出生率、死亡率、迁移率等。传统的人口预测方法往往基于统计模型或者简单的线性回归模型，但随着数据的复杂性和规模的增加，这些方法的准确性和预测能力逐渐受到限制。

近年来，神经网络在许多领域中取得了显著的成功，并且被广泛应用于预测和模式识别任务中。特别是，广义回归神经网络（Generalized Regression Neural Network, GRNN）作为一种非线性的回归模型，在数据建模和预测方面展现出了出色的性能。GRNN 具有强大的非线性建模能力和良好的泛化能

力,可以有效地捕捉输入变量之间的复杂关系,并输出准确的预测结果。本研究旨在基于广义回归神经网络,开发一种高精度的人口预测模型,以提供决策者和研究人员一个可靠的工具来预测人口趋势和变化。我们将使用历史人口数据作为输入,并训练 GRNN 模型来学习人口变化的模式和规律。通过结合多种相关的影响因素,如经济发展水平、社会政策变化和教育水平等,我们期望能够更准确地预测未来的人口趋势。

本研究的贡献包括:首先,我们将引入 GRNN 模型作为人口预测的新方法,通过对复杂的人口数据进行建模,提高预测的准确性和可靠性;其次,我们将结合多个相关因素,以更全面地考虑人口变化的驱动力,从而提高预测模型的解释能力;最后,我们将利用真实的人口数据集进行实证分析,并与传统方法进行对比,验证所提出模型的有效性和优越性。通过本研究,我们期望为人口预测提供一种新的建模方法和工具,为决策者提供更准确、可靠的人口信息,从而更好地指导社会规划和发展决策。

## 2 模型的建立与求解

### 2.1 数据的预处理与分析

基于上述数据进行预处理分析,为绘制全球 1950~2021 年的总人口随时间变化的趋势折线图,首先需要对所给数据进行 Excel 筛选处理,保留 Country 为“World”的数据,并利用 Origin 进行人口趋势图的绘制,如下图 1 所示。由图可知,全球 1950~2021 年总人口随时间变化的趋势折线图可得,全球整体人口的变化呈现逐年上升的趋势,且上升速度逐年加快。

### 2.2 广义回归神经网络 (GRNN) 模型建立

广义回归神经网络 (GRNN) 是一种基于梯度下降算法的神经网络模型,它最初由 Specht 在 1988 年提出。GRNN 是一种非线性模型,适用于分类和回归问题。与传统的神经网络模型不同,它使用了一种局部均值的概念,将数据分为两类<sup>[2]</sup>:核心数据和外围数据。GRNN 模型的训练过程步骤组成如下。

式 (1) 假设 GRNN 神经网络中的两个随机变量  $x$  和  $y$ , 其中,联合概率密度函数为  $f(x,y)$ , 且  $x$  的观测样本为  $X$ , 则条件均值为:

式 (2) 对于未知的概率密度函数  $f(x,y)$ , 可根据  $x$  和  $y$  的观测样本, 通过非参数进行估计得:

其中,变量  $X_i$  和  $Y_i$  分别表示随机变量  $x$  和  $y$  的

观测值,变量  $\delta$  表示平滑因子,变量表示样本数目,变量  $p$  表示随机变量  $x$  的维数。

由上式 (2) 可得式 (3)。

化简后可得式 (4)。

式 (4) 为人工神经网络的最终输出表达式。由上式可知, GRNN 神经网络的输出数据和训练样本的误差主要是由平滑因子决定的,所以, GRNN 神经网络的性能控制方式非常简便,只需通过调整平滑因子就可获得较好的性能<sup>[3]</sup>。

### (1) 模型的求解

根据以上模型,通过 Matlab 进行编程预测,首先分别对 1950~2021 年的中国、印度和全球的总人口数进行预测,并与实际人口数量进行比较,结果如下图所示。

由图 1950~2021 年的中国、印度和全球的实际人口与预测人口对比图可得,基于 TabNet 和广义回归神经网络的 CNN-BiLSTM-Attention 时序回归模型能较准确的对人口总数进行预测,则可进一步预测 2022~2100 年中国、印度和全球的人口总数,对结果进行可视化处理,

如下图所示<sup>[4]</sup>。由图可得,2022 到 2100 年,中国、印度和全球的总人口数量呈现缓慢上升的趋势,上升速度缓慢降低,其中,中国比印度的人口增长数更快逐渐趋近于平稳,则 2100 年底中国、印度和全球的总人口数,如下表所示。

## 3 结束语

未来研究方向可以从以下几个方面展开,以进一步改进基于广义回归神经网络的人口预测模型:

模型结构优化:进一步改进和优化广义回归神经网络的结构,包括网络层数、神经元数量、激活函数选择等。可以尝试引入更先进的神经网络模型,如循环神经网络 (RNN) 或长短期记忆网络 (LSTM),以捕捉时间序列数据中的长期依赖关系<sup>[5]</sup>。

数据特征提取:研究如何从原始的人口数据中提取更有价值的特征。可以尝试使用特征工程方法,结合领域知识和统计分析技术,提取与人口变化密切相关的特征,以提高模型的预测能力。

多源数据集成:将不同来源的数据集整合到人口预测模型中,例如社会经济数据、环境数据、医疗数据等<sup>[6]</sup>。通过融合多个数据源的信息,可以更全面地分析和预测人口的变化趋势,并提高模型的准确性。

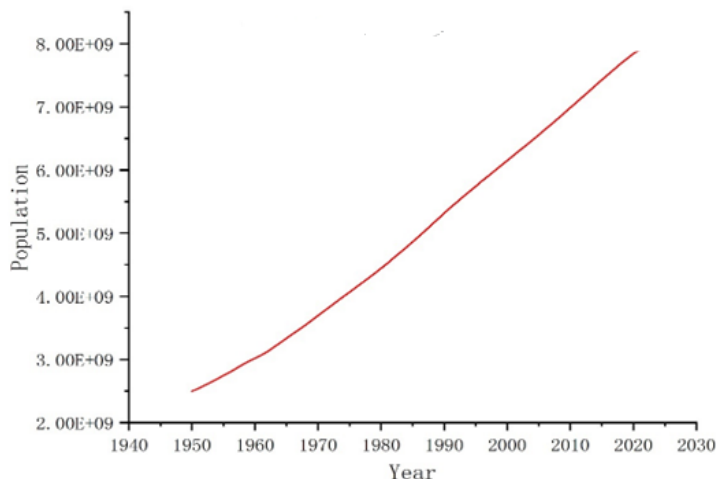


图1 1950-2021年世界总人口趋势折线图

$$\bar{Y} = E(y|X) = \frac{\int_{-\infty}^{+\infty} yf(X,y)dy}{\int_{-\infty}^{+\infty} f(X,y)dy} \tag{1}$$

$$\bar{f}(X,Y) = \frac{1}{(2\pi)^{\frac{(p+1)}{2}}\delta^{(p+1)}n} \cdot \sum_{i=1}^n \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\delta^2}\right] \exp\left[-\frac{Y-Y_i}{2\delta^2}\right] \tag{2}$$

$$\bar{Y} = \frac{\sum_{i=1}^n \left\{ \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\delta^2}\right] \int_{-\infty}^{+\infty} y \exp\left[-\frac{(y-Y_i)^2}{2\delta^2}\right] dy \right\}}{\sum_{i=1}^n \left\{ \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\delta^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{(y-Y_i)^2}{2\delta^2}\right] dy \right\}} \tag{3}$$

$$\bar{Y} = \frac{\sum_{i=1}^n \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\delta^2}\right]}{\sum_{i=1}^n \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\delta^2}\right]} \tag{4}$$

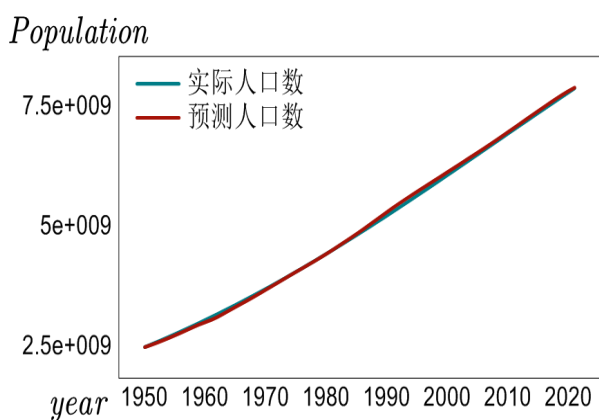


图2 全球实际人口与预测人口对比

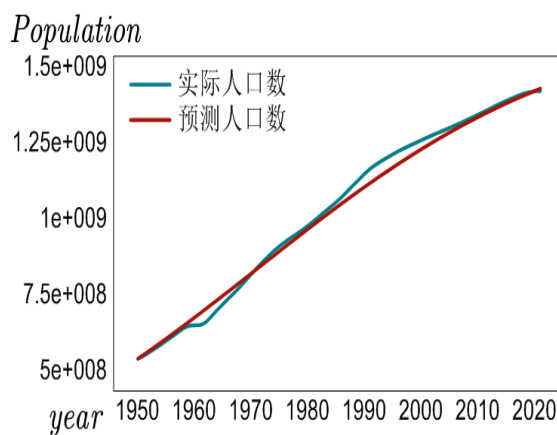


图3 中国实际人口与预测人口对比

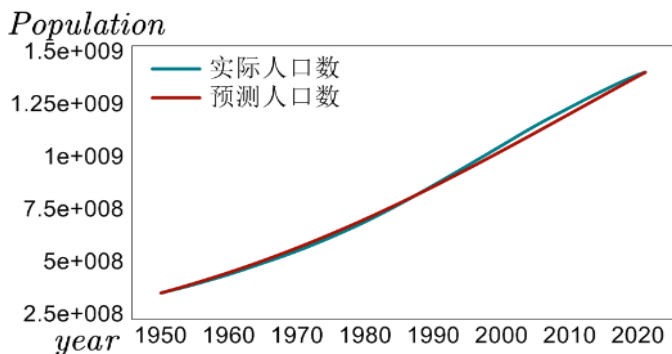


图4 印度 1950~2021 年实际人口与预测人口对比

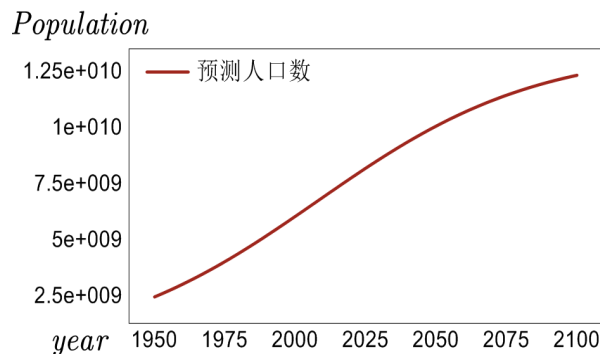


图5 全球 2022~2100 年人口预测

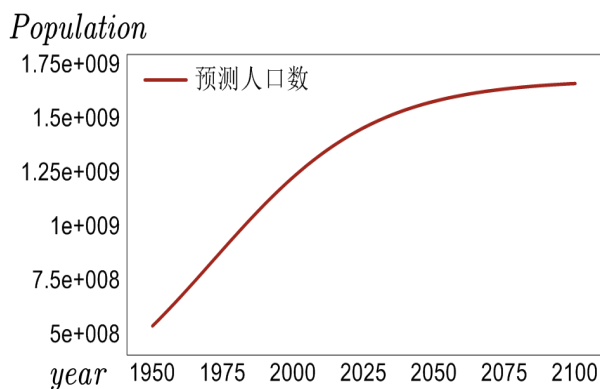


图6 中国 2022~2100 年人口预测

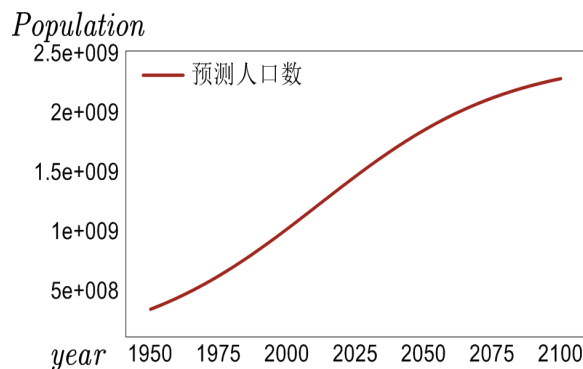


图7 印度 2022~2100 年人口预测

表1 2100 年底中国、印度和全球的总人口数

Country	Population
World	12368690121
China	1666878026
India	2297832195

预测不确定性分析：研究如何对人口预测结果的不确定性进行有效的分析和建模。考虑到人口预测中存在多种不确定因素，包括数据质量、模型误差等，可以引入概率建模或蒙特卡洛模拟等方法，量化预测结果的置信度和可靠性。

动态模型更新：建立动态的人口预测模型，能够实时或定期地更新预测结果，以适应人口变化的动态性。结合实时的人口数据和外部环境的变化，及时更新模型参数和预测结果，提高模型的时效性和适应性<sup>[7]</sup>。

跨地区预测：研究如何将基于广义回归神经网络的人口预测模型扩展到跨地区的预测任务中。考虑到不同地区的人口特征和发展差异，可以建立多个子模型或引入区域特征，以实现针对不同地区人口

变化的准确预测<sup>[8]</sup>。

综上所述，未来的研究可以从模型优化、数据特征提取、多源数据集成、不确定性分析、动态更新和跨地区预测等方面展开，以进一步提升基于广义回归神经网络的人口预测模型的准确性和应用价值。

### 参考文献

- [1] 安雨晴,杨宇,王莉.一种基于广义回归神经网络的裂纹扩展定量监测模型[J].航空科学技术,2023,34(03):40-48.
- [2] 张雪球,朱国庆.基于逻辑斯蒂回归模型的群体疏散行为研究[J].消防科学与技术,2014,33(03):330-332.
- [3] 员民,裴向东,乔钢柱,王莲.CNN-BiLSTM-Attention 模型在大规模计算系统故障预测的应用研究[J].计算机与数字工

- 程,2023,51(02):314-319+422.
- [4] 王怡,普运伟.基于 CNN-BiLSTM-Attention 融合神经网络的大气温度预测[J].中国水运(下半月),2023,23(01):25-27.
- [5] 朱益冬,陈玉明,卢俊文,曾念峰.融合 Logistic 回归与 Tabnet 模型的 P2P 网贷违约预测方法[J].厦门理工学院学报,2022,30(03):38-47.
- [6] 任志玲,王梓行.基于广义回归神经网络的煤矿带式输送机模型预测控制[J].辽宁工程技术大学学报(自然科学版),2023,42(01):92-98.
- [7] 王亦虹,李雅萱,田平野等.基于灰色-广义回归神经网络模型的城市群交通运输能力预测[J].重庆交通大学学报(自然科学版),2022,41(08):8-16.
- [8] 周敏,韩印,姚佼.基于广义回归神经网络的公交车运行时间预测模型[J].交通与运输(学术版),2013(02):75-78.

**版权声明:** ©2023 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<http://creativecommons.org/licenses/by/4.0/>



**OPEN ACCESS**