

# 基于灰色关联度分析和GRU神经网络的辽宁省GDP预测

扈文

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2021年10月15日; 录用日期: 2021年11月5日; 发布日期: 2021年11月17日

## 摘要

利用灰色关联分析方法分析了影响辽宁省GDP的宏观经济指标, 选取相关性较高的指标输入到门控循环单元神经网络中, 建立基于灰色关联分析的门控循环单元模型。首先利用灰色关联分析计算10个宏观经济指标与生产总值之间的相关度, 然后根据计算结果, 选取出5个关联度最大的指标建立灰色关联-GRU、灰色关联-LSTM和单一的GRU神经网络模型以及SVM模型对辽宁省生产总值进行预测, 将模型的预测结果进行比较, 结果表明基于灰色关联分析和GRU模型的预测精度较高。

## 关键词

灰色关联度分析, 门控循环单元, GDP

# GDP Prediction of Liaoning Province Based on Grey Correlation Analysis and GRU Neural Network

Wen Hu

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Oct. 15<sup>th</sup>, 2021; accepted: Nov. 5<sup>th</sup>, 2021; published: Nov. 17<sup>th</sup>, 2021

## Abstract

The macroeconomic indicators that affect GDP of Liaoning province are analyzed by using grey correlation analysis method, the indicators with high correlation with GDP are selected and input into the gated recurrent neural network, and the gated recurrent neural network model based on grey correlation analysis is established. Firstly, the correlation between 10 macroeconomic indicators and GDP is calculated by grey correlation analysis, then five indexes with strong correlation

from several macroeconomic indicators are selected to establish GRU neural network based on grey correlation analysis, LSTM neural network based on grey correlation analysis, GRU neural network and SVM to predict the GDP of Liaoning Province. Comparing the prediction results of these models, the results show that the GRU neural network based on grey correlation analysis has high prediction accuracy.

## Keywords

Grey Correlation Analysis, Gated Recurrent Neural Network, GDP

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

## 1. 引言

国内生产总值(Gross Domestic Product), 是衡量一个国家或地区经济发展情况的指标之一。新中国成立后, 辽宁是新中国工业的摇篮, 为新中国贡献“1000多个全国第一”, 辽宁省的经济在稳步增长。但是在2015年和2016年, 辽宁省GDP出现大幅度的负增长, 自2017年以来, 虽然辽宁GDP仍在增长, 但经济速度发展缓慢。所以对辽宁省GDP的发展趋势做出尽可能准确地预测, 对其制定未来的有关经济社会发展政策会有重要的参考价值。

## 2. 相关研究综述

目前已有许多学者使用不同方法对GDP数据进行预测。王佳佳[1]选取1978~2019年安徽省地区生产总值进行分析, 建立了ARIMA(2,1,3)模型, 并对2020~2022年的GDP做出预测。祖培福等[2]运用2013~2018年黑龙江省GDP数据建立了优化初始条件的GM(1,1)模型, 并对2019~2022年黑龙江省GDP做出了预测。罗志丹等[3]选取了1992年第1季度至2018年第4季度中国GDP季度数据, 并应用EEMD方法对数据进行分解, 然后对分解后的数据进行重构, 建立粒子群优化算法优化的反向传播神经网络和ARIMA模型, 结果发现EEMD-PSOBPNN-ARIMA模型的预测效果更好。潘典雅[4]对1993~2017年吉林省的生产总值进行分析, 发现序列自相关系数和偏相关系数均为拖尾, 因此建立了ARIMA模型, 在多个ARIMA模型中比较它们的预测结果, 并结合AIC准则, 发现ARIMA(2,1,1)模型的预测精度较高, 因此采用ARIMA(2,1,1)对未来三年吉林省的生产总值做出预测。朱青, 周石鹏[5]对1980~2018年的GDP增长进行预测, 由于数据样本较少, 且样本值变化大, 不适合使用传统的时序模型进行预测, 因此提出一种新的方法, 首先利用随机森林对影响GDP增长的变量进行重要性排序, 然后选取重要变量。再运用LSTM网络, 根据选取的变量对GDP增长进行预测分析, 将预测结果与传统模型比较, 结果表明基于随机森林的LSTM模型预测效果更好。Hao Lv [6]采用自回归综合移动平均模型(ARIMA)、BP神经网络、支持向量机(SVM)三种机器学习方法对1998~2017年中美GDP进行了分析、建模和预测。用相对误差来评估所提出模型的准确性, 通过对三种方法的比较, 发现BP神经网络最适合用于GDP预测分析。

## 3. 灰色关联度分析与GRU神经网络

### 3.1. 灰色关联度分析

灰色关联分析方法, 是根据几个因素之间发展趋势的相似程度, 来衡量因素间关联程度的一种方法。

下面介绍灰色关联度分析的计算步骤。

1) 建立各相关指标的原始数据矩阵

$$x_i = (x_i(1), x_i(2), \dots, x_i(k), \dots) \quad (1)$$

2) 对参考数列和比较数列进行无量纲化处理，本文采用均值化处理。

3) 计算灰色关联系数。

各比较数列与参考数列在各个时刻的关联系数计算公式如下：

$$\zeta_i = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|} \quad (2)$$

其中  $\rho$  是一个可调节的系数，取值为(0,1)，目的是为了调节输出结果的差距大小，本文取  $\rho = 0.5$ 。

4) 计算灰色关联度

由于上一步计算出的比较数列与参考数列的关联系数的值不止一个，所以需要计算其平均值作为关联程度的数量表示。计算公式如下

$$r_i = \frac{1}{N} \sum_{k=1}^N \zeta_i(k) \quad (3)$$

若  $r_i$  的值越接近 1，则相关性越好。

### 3.2. GRU 神经网络

门控循环单元(Gated Recurrent Units)是解决循环神经网络(RNN)梯度消失问题的另外一种方法，它是通过门来控制信息的流动。是在长短期记忆网络(LSTM)的基础上形成的，与 LSTM 网络不同的是，GRU 只使用了两个门，把输入门和遗忘门合并成更新门。更新门的计算公式为：

$$Z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \quad (4)$$

其中， $Z_t$  表示更新门的输出值， $x_t$  为  $t$  时刻的输入值， $h_{t-1}$  为  $t-1$  时刻的输出值， $\sigma$  为 sigmoid 函数， $W_{xz}$  与  $W_{hz}$  为权重， $b_z$  为偏置。更新门决定了之前的记忆值进入当前值的比例。另外一个门是重置门，定义如下：

$$r_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \quad (5)$$

记忆单元的值定义为：

$$C_t = \tanh(W_{xc}x_t + W_{rc}(h_{t-1} \circ r_t) + b_c) \quad (6)$$

它由上一个时刻的状态值及当前输入值共同决定。隐含层的状态值定义为：

$$h_t = (1 - Z_t) \circ C_t + Z_t \circ h_{t-1} \quad (7)$$

其中， $r_t$  表示重置门， $\tanh$  为激活函数， $C_t$  为过程量。

GRU 算法中的门控结构同样拥有可以学习长序列的特征，而且比 LSTM 算法更简单。

## 4. 实证分析

### 4.1. 数据选取

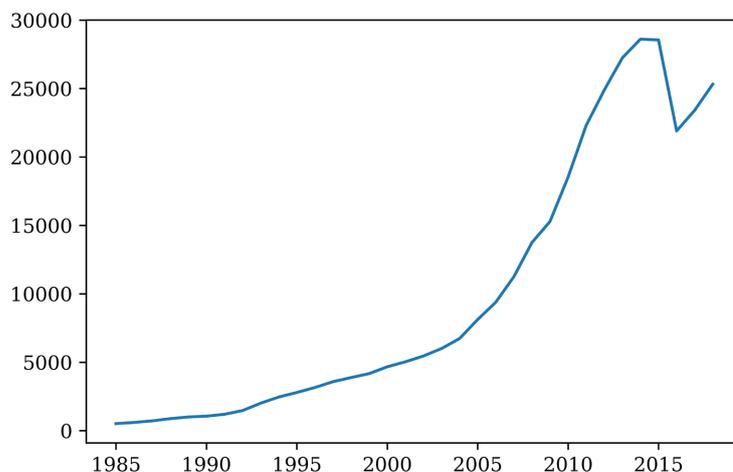
选取 1975~2018 年的 34 条数据，包括生产总值，第一产业，第二产业，第三产业，人均生产总值，

居民消费价格指数，城市居民消费价格，农村居民消费价格，商品零售价格指数，公共财政预算收入，年末总人口数 11 个指标。数据如下表 1 所示。其中 1985~2014 年共 30 条数据作为训练样本，2015~2018 年共 4 条作为测试样本。

**Table 1.** GDP of Liaoning Province  
**表 1.** 辽宁省 GDP

年份/年	GDP/亿元	年份/年	GDP/亿元	年份/年	GDP/亿元
1985	518.6	1997	3582.5	2009	15,288.67
1986	605.3	1998	3881.7	2010	18,528.61
1987	719.1	1999	4171.7	2011	22,301.47
1988	881	2000	4669.1	2012	24,882.59
1989	1003.8	2001	5033.1	2013	27,246.23
1990	1062.7	2002	5458.2	2014	28,612.34
1991	1200.1	2003	6002.5	2015	28,555.57
1992	1473	2004	6736.12	2016	21,896.19
1993	2010.8	2005	8123.13	2017	23,409.24
1994	2461.8	2006	9388.89	2018	25,315.35
1995	2793.4	2007	11,248.97		
1996	3157.7	2008	13,745.32		

以上数据均来源于辽宁省统计年鉴。在图 1 中可以看生产总值变化幅度较大，这是因为二十世纪八十年代时生产技术不高，但随着科技的发展，二十一世纪我国经济高速发展，所以在 2000~2015 年，辽宁省 GDP 总体上增长较快，而 2015 年后辽宁的 GDP 出现下滑趋势，甚至出现负增长的情况。



**Figure 1.** GDP of Liaoning province  
**图 1.** 辽宁省 GDP

## 4.2. 数据预处理

本文使用的软件为 Python，在使用数据之前先将数据进行归一化处理，消除量纲的影响，本文采用“最大最小值”标准化方法，即  $x' = \frac{x - x_{\max}}{x_{\max} - x_{\min}}$ ， $x$  为原始数据， $x'$  为标准化的数据， $x_{\max}$  为原始数据的最大值， $x_{\min}$  为原始数据的最小值。使用 Python 中的 MinMaxScaler 命令，将数据转化为(0,1)之间的数。

## 4.3. 模型建立

### 4.3.1. 基于灰色关联分析的 GRU 模型

将归一化后的数据进行灰色关联分析，其中生产总值为参照序列，而第一产业，第二产业，第三产业，人均生产总值，居民消费价格指数，城市居民消费价格，农村居民消费价格，商品零售价格指数，公共财政预算收入，年末总人口数为比较序列，计算生产总值与每个影响因素之间的关联度大小并进行排序，结果如表 2 所示。

通过以上计算，选取与生产总值相关度在 0.8 以上的指标作为训练指标，有人均生产总值、第二产业、第三产业、第一产业、公共财政收入预算。将以上指标输入到 GRU 神经网络中，将人均生产总值、第二产业、第三产业、第一产业、公共财政收入预算作为输入，GDP 作为输出。神经网络的训练损失函数为平方绝对误差(mean absolute error, MAE)，考虑到 Adam 优化器计算高效，所以选择 Adam 作为优化器，设置迭代次数为 1000 次。用前一年的特征指标预测后一年的生产总值。

**Table 2.** Grey correlation degree and sorting  
**表 2.** 灰色关联度排序

指标	关联度大小
人均生产总值	0.993557
第二产业	0.955088
第三产业	0.952057
第一产业	0.929751
公共财政收入预算	0.923374
年末总人口数	0.764163
商品零售价格指数	0.665279
农村居民消费价格指数	0.659497
居民消费价格指数	0.640659
城市居民消费价格指数	0.632381

### 4.3.2. 对比模型

1) LSTM 网络善于处理和预测较长的时间序列，因此选择建立基于灰色关联分析的 LSTM 模型作为对比模型，模型中选用 ReLU 作为激活函数，并添加 Dropout 层防止过拟合，迭代次数同样为 1000 次。

2) 本文还建立了单一的 GRU 网络作为对比模型，与基于灰色关联分析的 GRU 网络相同，选择平方绝对误差为损失函数，Adam 作为优化器，迭代次数为 1000 次。

3) 除此之外, 还选择了支持向量机模型作为对比模型, 经过多次实验, 最终将参数设置为  $C = 1$ ,  $\gamma = 10$ , 核函数选择 rbf 核函数。

#### 4.4. 模型对比

为了判断预测的准确度, 本文采用均方根误差(RMSE)、均方误差(MSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)对预测结果进行评价。均方误差表示当预测值与真实值完全吻合时均方误差为 0, 此时为完美模型, 误差越大, 该值越大, 计算公式如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_{pred,i} - X_{real,i})^2 \quad (8)$$

平均绝对误差能更好地反映预测值误差的实际情况, 误差越大, 该值越大, 计算公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_{pred,i} - X_{real,i}| \quad (9)$$

多种模型的训练和预测结果如表 3 所示, 实际值代表真实的辽宁省 GDP 数据。从实际值与预测值来看, 可以发现基于灰色关联分析的 GRU 神经网络模型预测效果很好, 除了 2016 年 GDP 的实际值显著下降, 而模型没有学习到如此大幅度的下降, 所以预测值下降的较少才导致预测误差较大, 其余都可以得到较为精准的预测值。基于灰色关联分析的 LSTM 神经网络模型与单一的 GRU 神经网络模型的预测值都会比实际值偏高, 而 SVM 模型的预测值远远低于实际值, 特别是在 2016~2018 年之间, 实际值是处在上升趋势, 而 SVM 模型的预测值一直在下降, 所以相比较其它模型 SVM 模型的预测效果较差。从评价指标 MSE 和 MAE 的对比中发现将数据进行灰色关联分析后在使用 GRU 神经网络模型比直接使用 GRU 神经网络模型预测误差明显降低, 而基于灰色关联分析的 LSTM 神经网络预测的误差较大, 单一的 GRU 神经网络模型的预测结果较稳定, 但是总体误差偏大, SVM 模型的预测效果一般。通过上述实证分析, 基于灰色关联分析的 GRU 神经网络能较好的对辽宁省 GDP 进行预测。

**Table 3.** Comparison of prediction results

**表 3.** 预测结果对比

年份	实际值	灰色关联-GRU	灰色关联-LSTM	GRU	SVM
2015	28,555.57	34,505.33	37,055.53	35,939.83	22,690.21
2016	21,896.19	31,654.98	34,308.99	33,077.83	12,605.37
2017	23,409.24	26,349.53	27,741.95	27,340.24	12,269.44
2018	25,315.35	28,237.66	29,821.95	29,353.25	12,092.86
	MSE	0.046822	0.084069	0.066934	0.132926
	MAE	0.438129	0.514545	0.485929	0.593015

### 5. 总结

本文建立了基于灰色关联分析的 GRU 神经网络、基于灰色关联分析的 LSTM 神经网络、单一的 GRU 神经网络以及 SVM 模型对辽宁省 GDP 进行预测。常用的 GRU 神经网络对于辽宁省的 GDP 预测较为准确, 其中基于灰色关联分析的 GRU 神经网络模型对辽宁省 GDP 的预测精度高于单一的 GRU 神经网络模型。这一结果表明, 在影响辽宁省 GDP 的多个指标中, 选取相关度高的指标来进行预测, 会提升预测

性能。而基于灰色关联的 GRU 神经网络预测效果好于基于灰色关联分析的 LSTM 神经网络, 说明 GRU 神经网络在预测短期的 GDP 方面性能优于 LSTM 模型。通过 SVM 模型的实验结果可以发现, GRU 神经网络比传统的机器学习方法对非线性数据的预测更有优势。综上分析, 使用基于灰色关联分析的 GRU 神经网络模型进行辽宁省 GDP 预测具有一定的实用性、可行性, 可为决策者提供有效参考。本文只采用了一些宏观经济数据建立模型进行预测, 而影响辽宁省 GDP 变化的还包括很多, 如经济运行机制、国家政策、经济结构等, 所以模型中难免存在不足, 还需要进一步改善, 提高神经网络预测辽宁省 GDP 的精度。

### 参考文献

- [1] 王佳佳. 基于 ARIMA 模型的安徽省 GDP 的分析与预测[J]. 枣庄学院学报, 2020, 37(5): 38-44.
- [2] 祖培福, 王嘉婷, 卢含. 基于初始条件优化的 GM(1.1)模型在黑龙江省 GDP 预测中的应用[J]. 商业经济, 2020(2): 8-9+64.
- [3] 罗志丹, 刘英. 中国季度 GDP 预测模型比较[J]. 统计与决策, 2020, 36(5): 33-37.
- [4] 潘典雅. 基于 ARIMA 模型的吉林省 GDP 分析及预测[J]. 中国集体经济, 2021(27): 15-16.
- [5] 朱青, 周石鹏. 基于 LSTM 模型的国民经济 GDP 增长预测建模研究[J]. 经济研究导刊, 2021(19): 5-9+37.
- [6] Lv, H. (2020) Chinese and American GDP Forecasts Based on Machine Learning. *World Scientific Research Journal*, 6, 95-104.