

基于 RGCVAE 的测井曲线重构方法

韩建^{1,2}, 陈着^{1,2}, 王业统³, 曹志民^{1,2}, 邓宇^{1,2}

(1. 东北石油大学 三亚海洋油气研究院, 海南 三亚 572000;

2. 东北石油大学 物理与电子工程学院, 黑龙江 大庆 163318;

3. 海南科技职业大学 虚拟现实技术与系统海南省工程研究中心, 海南 海口 571126)

摘要:在实际测井过程中,测井曲线的质量常常受到仪器故障和环境因素的影响,导致测井数据出现缺失。该文提出了一种基于 RGCVAE 的测井曲线重构方法,并结合大庆油田古工业区和金工业区的实际测井数据,分别进行了同井间和异井间的缺失数据重构实验。通过与随机森林、RNN 和 LSTM 网络的实验结果进行对比分析,结果表明,RGCVAE 模型在预测精度方面表现较好。在同井实验中,两口井重构后的声波时差曲线原始曲线的相关性分别达到了 90.94% 和 88.60%;在异井实验中,两口井重构后的声波时差曲线与原始曲线的相关性分别为 87.85% 和 85.71%。

关键词:循环格兰杰变分编码器;测井曲线;重构方法;声波时差曲线

中图分类号:TP391

文章编号:1000-0682(2025)05-0087-05

文献标识码:A

DOI:10.19950/j.cnki.CN61-1121/TH.2025.05.016

Reconstruction method of logging curves based on RGCVAE

HAN Jian^{1,2}, CHEN Zhuo^{1,2}, WANG Yetong³, CAO Zhimin^{1,2}, DENG Yu^{1,2}

(1. Northeast Petroleum University Sanya Institute of Marine Oil and Gas Research, Hainan Sanya 572000, China; 2. School of Physics and Electronic Engineering, Northeast Petroleum University, Heilongjiang Daqing 163318, China; 3. Hainan Vocational University of Science and Technology Virtual Reality Technology and System Hainan Provincial Engineering Research Center, Hainan Haikou 571126, China)

Abstract:In the actual logging process, the quality of logging curves is often compromised by instrument malfunctions and environmental factors, leading to missing logging data. This paper introduces a logging curves reconstruction method based on recurrent granger causality variational autoencoder. By utilizing actual logging data from the ancient and gold industrial areas of Daqing Oilfield, we conducted missing data reconstruction experiments for both intra-well and inter-well scenarios. Through comparative analysis with experimental results obtained from random forest, recurrent neural network, and long short-term memory networks, it was found that the recurrent granger causality variational autoencoder model outperforms other methods in terms of prediction accuracy. In the intra-well experiment, the correlation of the reconstructed acoustic moveout curve and the original curve of the two wells reached 90.94% and 88.60%, respectively. In the inter-well experiment, the correlation between the reconstructed acoustic time difference curve and the original curve of the two wells is 87.85% and 85.71%.

Keywords:recurrent granger causality variational autoencoder; logging curve; reconstruction method; acoustic time difference curve

0 引言

实际测井过程中,由于仪器故障、井壁不稳定、地层异质性等因素可能会导致部分深度范围内的声波时差测井曲线无法被准确记录,甚至出现完全缺

失的情况。重新测井操作既费时又成本高昂,并且在复杂地质条件下,可能无法完全恢复所需的数据^[1-3]。因此,学者对测井曲线进行重构研究。

传统的测井曲线重构方法主要包括经验公式法和线性回归法^[4]。经验公式法依赖于岩性与测井数据之间的经验性关系,线性模型法则通过建立测井曲线之间的线性关系来实现重构,对复杂地层预测时误差较大。近年来,针对传统模型存在的局限性,机器学习技术被引入到测井曲线重构实验中。

收稿日期:2025-03-03

第一作者:韩建(1976—),男,汉族,黑龙江大庆人,博士,教授,研究方向为油井信号检测,机器学习,模式识别。

例如,文献[5]提出多元非线性回归计算模型,将复杂的非线性关系映射为多个线性关系重构测井曲线。文献[6]等人提出基于遗传神经网络优化方法,可以有效地克服易陷入局部最小值的缺点。文献[7]提出串级双向长短时记忆神经网络,可以充分考虑缺失点上下关联信息。

虽然这些预测方法在重构能力上表现出色,但是没有利用测井曲线之间的因果关系。该文提出了一种基于循环格兰杰变分编码器^[8](Recurrent Granger Causality Variational Autoencoder, RGCVAE)的测井曲线重构方法。该模型通过有效捕捉不同变量之间的格兰杰因果关系,利用动态因果结构重构缺失曲线。通过对大庆油田古工业区和金工业区的实验验证,RGCVAE 模型能够作为一种基准的测井曲线的重构方法。

1 原理与方法

1.1 格兰杰因果关系

格兰杰因果关系是一种统计假设检验方法,用于判断一个时间序列是否对另一个时间序列具有预测能力,通常用于检验因果关系的方向性。在经济学、金融学、气候学、神经科学等多个领域中广泛应用^[9-11]。图 1 展示了测井曲线间的格兰杰因果关系。格兰杰因果关系的基本思想是,如果一个时间序列 X 的过去值有助于预测另一个时间序列 Y 的当前值,则可以认为 X 格兰杰“引起” Y ^[12]。

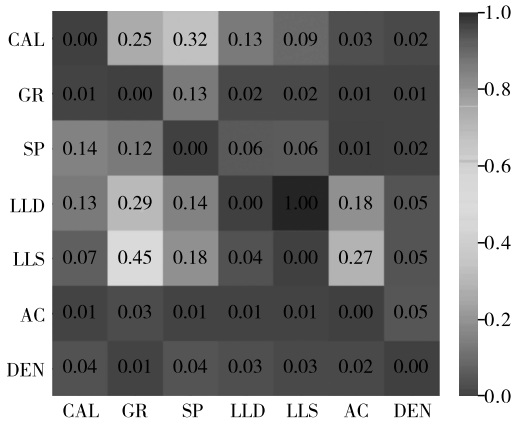


图 1 格兰杰因果关系矩阵热图

格兰杰因果关系检验通常使用向量自回归模型来进行。具体步骤如下:

假设有两个时间序列,首先对每个变量建立自回归模型。假设选择的滞后阶数为 p ,则自回归模型可以表示为:

$$Y_t = \alpha_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \delta_t \quad (1)$$

$$X_t = \gamma_0 + \sum_{i=1}^p \delta_i X_{t-i} + \eta_t \quad (2)$$

式中: α_0 和 γ_0 为常数项; β_i 和 δ_i 为回归系数; δ_t 和 η_t 为误差项。

再将两个时间序列一起考虑,建立一个向量自回归模型:

$$Y_t = \alpha_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \sum_{i=1}^p \lambda_i X_{t-i} + \delta_t \quad (3)$$

$$X_t = \gamma_0 + \sum_{i=1}^p \delta_i X_{t-i} + \sum_{i=1}^p \mu_i Y_{t-i} + \eta_t \quad (4)$$

式中: λ_i 是源时间序列 X_t 的滞后值对目标时间序列的影响; μ_i 是目标时间序列 Y_t 滞后值对源时间序列的影响。

如果源时间序列 X_t 对目标时间序列 Y_t 没有格兰杰因果关系,即:

$$\lambda_1 = \lambda_2 = \dots = \lambda_p = 0 \quad (5)$$

如果源时间序列 X_t 对目标时间序列 Y_t 存在格兰杰因果关系,即至少有一个滞后项对目标变量的预测具有显著影响 $\lambda_i \neq 0$,则通过 F 检验评估两者之间的显著性差异。如果引入滞后项显著提高了模型的预测能力,则拒绝零假设,进而证明源时间序列对目标时间序列具有格兰杰因果关系。

1.2 循环格兰杰变分编码器

RGCVAE 由递归格兰杰因果关系重构模块和基于域敏感的格兰杰因果预测模块组成,其结构如图 2 所示。

1.2.1 递归格兰杰因果关系重构模块

递归格兰杰因果关系重构模块旨在从时间序列数据中重建格兰杰因果关系结构。每个滞后的因果结构通过以下递归关系得到:

$$\begin{cases} \mathbf{A}_1^* = f_1(z_{t-1}^*, z_{t-2}^*, \dots, z_{t-k}^*, \alpha^*; \theta_1) \\ \mathbf{A}_2^* = f_2(\mathbf{A}_1^*, z_{t-1}^*, z_{t-2}^*, \dots, z_{t-k}^*, \alpha^*; \theta_2) \\ \dots \\ \mathbf{A}_k^* = f_k(\mathbf{A}_1^*, \mathbf{A}_2^*, \dots, \mathbf{A}_{k-1}^*, z_{t-1}^*, z_{t-2}^*, \dots, z_{t-k}^*, \alpha^*; \theta_k) \end{cases} \quad (6)$$

式中: f_i 由多层感知机网络(Multilayer Perceptron, MLP)组成;* 代表数据来自源域或目标域; $z_{t-1}^*, z_{t-2}^*, \dots, z_{t-k}^*$ 是历史时间序列数据; α^* 是领域特定的潜在变量; θ_i 是每个多层感知机网络的可训练参数; \mathbf{A}_i^* 是因果关系矩阵。

1.2.2 基于域敏感的格兰杰因果预测模块

在基于域敏感的格兰杰因果预测模块中将源域和目标域的历史时序数据结合,通过捕捉滞内和滞间格兰杰因果关系预测目标领域未来时刻数据。

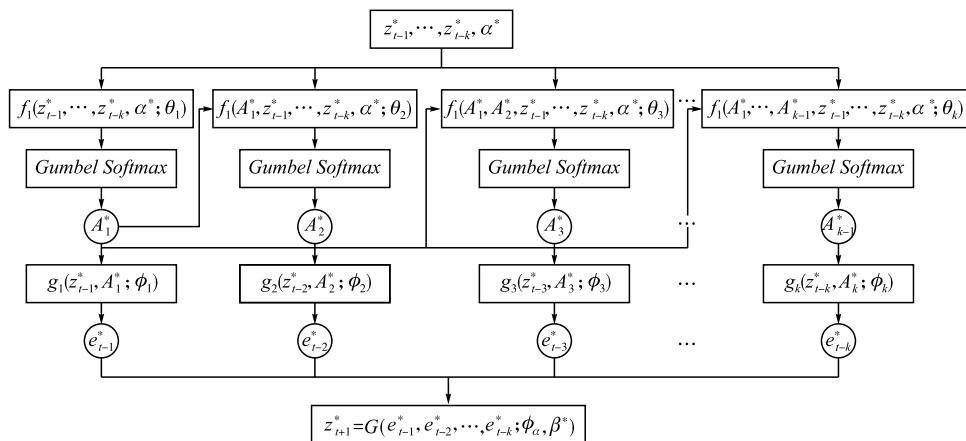


图2 RGCVAE 结构图

在滞内预测过程中,模型根据每个滞后的因果结构和历史数据计算每个子结构对最终预测的影响 $e_{t-1}, e_{t-2}, \dots, e_{t-k}$, 即:

$$\begin{cases} e_{t-1} = g_1(z_{t-1}^*, A_1^*, \phi_1) \\ e_{t-2} = g_2(z_{t-2}^*, A_2^*, \phi_2) \\ \dots \\ e_{t-k} = g_k(z_{t-k}^*, A_k^*, \phi_k) \end{cases} \quad (7)$$

式中: g_i 是每个滞后步骤的因果预测函数; ϕ_i 是可学习的参数。

在滞间预测阶段,模型聚合所有时间滞后的贡

献,来预测目标变量的未来值 z_{t+1}^* :

$$z_{t+1}^* = G(e_{t-1}^*, e_{t-2}^*, \dots, e_{t-k}^*; \phi_\alpha, \beta^*) \quad (8)$$

式中: $G(\cdot)$ 是滞间预测函数; $\Phi = (\phi_\alpha, \phi_1, \dots, \phi_k, \beta)$ 是基于格兰杰因果关系的时滞预测中可训练参数。

2 实验与结果分析

为了测试 RGCVAE 模型重构测井曲线的性能,该文进行了同井实验和异井实验,原理如图 3 所示。

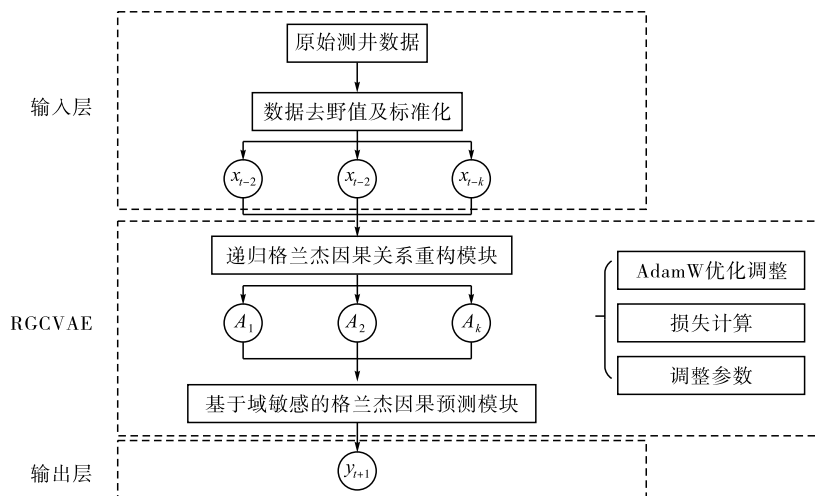


图3 RGCVAE 模型重构实验原理图

2.1 实验数据

数据集选自大庆油田古工业区 A1 井、A2 井、A3 井、A4 井和金工业区 B1 井、B2 井实测数据。A1 深度范围是 1200 m 到 2206.5 m、A2 井的深度范围是 1215 m 到 2219.5 m、A3 井的深度范围是 1230 m 到 1404.875 m、A4 井的深度范围是 1550.625 m 到 1925.5 m、B1 井的深度范围是 1624.375 m 到 1999.25 m、B2 井的深度范围是 1685.5 m 到

2060.375 m。每口井的采样间隔为 0.125 m,且均包含 6 条测井曲线:自然伽马(Gamma Ray, GR)、声波时差(Acoustic, AC)、密度(Density, DEN)、深侧向(Deep Laterolog, LLD)、井径(Caliper, CAL)以及自然电位(Self Potential, SP)。

2.2 数据预处理

在实测过程中,如果井壁存在不规则的形态,测井工具可能会受到影响,导致测量数据不稳定,产生

异常值。因此需要对测井数据做去异常值处理,如图 4 所示。

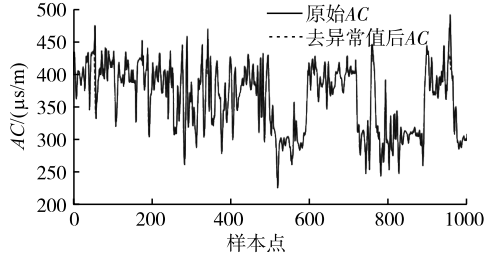


图 4 去异常值示意图

如果某个数据点的值超出均值的 3 倍标准差范围,则该数据点被认为是异常值,即:

$$|x_i - \mu| > 3\sigma \quad (9)$$

式中: x_i 是数据点; μ 是均值; σ 是标准差。

为了消除测井曲线范围差异较大的影响,对去除异常值后的测井曲线进行标准化处理:

$$s_i = \frac{x_i - \mu}{\sigma} \quad (10)$$

式中: s_i 为标准化后的测井曲线数据。

2.3 参数配置

该文中 RGCVAE 模型的递归格兰杰因果关系重构模块由一个 MLP 网络、批归一化层、丢弃层组成,基于域敏感的格兰杰因果预测模块由一个 MLP 网络、ReLU 激活层、批归一化层和丢弃层组成。模型的初始学习率设置为 0.05。数据的批量大小为 512,最大时间滞后为 5。

评价指标使用均方误差 (Mean Squared Error, MSE)、均方根误差 (Root Mean Squared Error, RMSE)、平均绝对值误差 (Mean Absolute Error, MAE) 以及皮尔逊相关系数 (Pearson Correlation Coefficient, PCCs), 即:

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - x'_i)^2 \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x'_i)^2} \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - x'_i| \quad (13)$$

$$PCCs = \frac{\text{cov}(x_i, x'_i)}{\sigma_{x_i} \cdot \sigma_{x'_i}} \quad (14)$$

式中: x_i 是真实值; x'_i 是预测值; N 是样本个数。

2.4 同井 AC 曲线复原

在同井实验中,选择 A1 井和 A2 井的 70% 数据作为训练集,30% 数据作为验证集。为了模拟 AC 曲线缺失的情况,完全删除 A3 井和 A4 井的 AC 曲线,并利用其他完整的测井曲线对 AC 曲线进行重

构,如图 5、6 所示。实线是原始值,虚线是预测值。

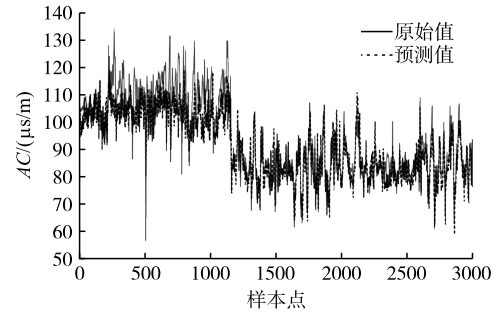


图 5 A3 井 AC 曲线重构效果图

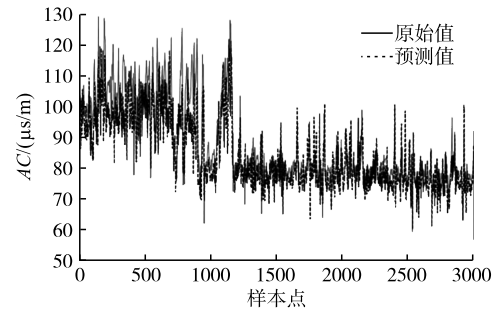


图 6 A4 井 AC 曲线重构效果图

表 1 A3 井 AC 曲线重构定量结果

方法	MSE	RMSE	MAE	PCCs
随机森林	164.003 1	12.806 4	8.676 8	0.760 2
RNN	52.228 2	7.226 9	5.352 3	0.870 3
LSTM	41.890 3	6.4723	4.816 6	0.900 7
RGCVAE	36.810 3	6.067 2	4.429 5	0.909 4

表 2 A4 井 AC 曲线重构定量结果

方法	MSE	RMSE	MAE	PCCs
随机森林	110.048 0	10.490 4	7.236 0	0.751 6
RNN	61.218 7	7.824 2	5.443 5	0.843 9
LSTM	52.220 0	7.226 3	5.352 0	0.870 4
RGCVAE	52.655 6	7.256 4	5.217 4	0.886 0

表 3 B1 井 AC 曲线重构定量结果

方法	MSE	RMSE	MAE	PCCs
随机森林	289.105 0	17.003 1	14.533 7	0.702 2
RNN	81.236 5	9.013 1	6.490 6	0.873 9
LSTM	38.186 5	6.179 5	4.123 0	0.923 3
RGCVAE	58.996 6	7.680 9	5.344 4	0.878 5

表 4 B2 井 AC 曲线重构定量结果

方法	MSE	RMSE	MAE	PCCs
随机森林	75.655 2	8.698 0	5.971 3	0.773 0
RNN	93.638 7	9.676 7	6.868 6	0.6934
LSTM	59.162 3	7.691 7	5.860 5	0.803 6
RGCVAE	49.880 4	7.063 2	5.603 8	0.857 1

通过图 5 和图 6 中原始值和预测值两条曲线的对比,可以看出预测值在大多数情况下较好地拟合了原始值,特别是在低频波动的情况下,两者之间的

差异相对较小。

将随机森林、RNN 与 LSTM 的实验结果与 RGCVAE 模型对比,表 1 为使用不同方法对 A3 井 AC 曲线重构的定量结果。可以看出 RGCVAE 模型的重构误差最小为 36.810 3 (MSE)、6.067 2 ($RMSE$)、4.429 5 (MAE),预测的 AC 曲线与原始曲线的相关系数最大为 90.94%。表 2 为 A4 井 AC 曲线重构的定量结果。可以看出 RGCVAE 模型预测的 AC 曲线与原始曲线的相关系数最大为 88.60%。

2.4 异井 AC 曲线复原

在异井实验中,选择 A1 井和 A2 井的 70% 数据作为训练集,30% 数据作为验证集。B1 和 B2 井作为测试集,并将 AC 曲线完全删除,模拟重构情况。异井的重构效果如图 7 和图 8 所示,实线代表原始值,虚线代表预测值。

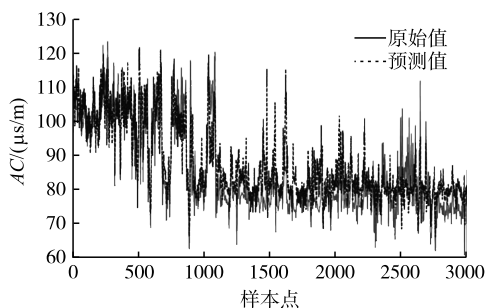


图 7 B1 井 AC 曲线重构效果图

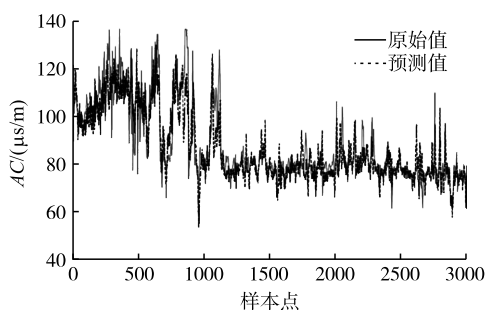


图 8 B2 井 AC 曲线重构效果图

从图 7 和图 8 中可以看出,预测曲线与原始数据在波动趋势上具有较高的相似性,表明模型能够有效地捕捉到数据的主要变化模式。然而,在一些特定的局部区域,预测值与原始值之间存在一定的偏差,表明模型在这些细节变化上的拟合精度有所欠缺。

表 3 展示了使用不同方法对 B1 井 AC 曲线重构的定量评估结果。结果表明,RGCVAE 模型的 MSE 、 $RMSE$ 、 MAE 分别为 58.996 6、7.680 9 和 5.344 4,均明显优于随机森林模型。此外,RGCVAE 的 $PCCs$ 为 0.878 5,虽然略逊色于 LSTM,但仍显示出良好的相关性。表 4 表明,RGCVAE 模型的

预测精度均高于其他方法,并且 $PCCs$ 为 0.857 1。

3 结论

该文从测井曲线间的格兰杰因果关系角度进行预测,使用 RGCVAE 模型以大庆油田两个工业区内油井为实验对象,进行了同井和异井 AC 曲线的重构实验。通过与随机森林、RNN 以及 LSTM 方法比较,实验结果表明,RGCVAE 模型可以有效地重构测井曲线。然而,异井预测的精度较低于同井预测,说明 RGCVAE 模型在处理异井数据时仍存在一定的局限性。未来的工作将重点关注减小异井预测的误差,增强模型的鲁棒性。

参考文献:

- [1] 王锦涛,文晓涛,何易龙,等.基于 CNN-GRU 神经网络的测井曲线预测方法[J].石油物探,2022,61(02):276-285.
- [2] 张家臣,邓金根,谭强,等.基于 XGBoost 的测井曲线重构方法[J].石油地球物理勘探,2022,57(03):697-705.
- [3] 尚福华,卢玉莹,曹茂俊.基于改进 LSTM 神经网络的测井曲线重构方法[J].计算机技术与发展,2022,32(06):198-202.
- [4] 翟晓岩,高刚,李勇根,等.融合注意力机制的二维卷积神经网络测井曲线重构方法[J].石油地球物理勘探,2023,58(05):1031-1041.
- [5] 王俊瑞,梁力文,邓强,等.基于多元回归模型重构测井曲线的方法研究及应用[J].岩性油气藏,2016,28(03):113-120.
- [6] 金永吉,张强,王毛毛.基于遗传神经网络算法的测井曲线重构技术[J].地球物理学进展,2021,36(03):1082-1087.
- [7] 周伟,赵海航,蒋云凤,等.基于串级双向长短时记忆神经网络的测井数据重构[J].石油地球物理勘探,2022,57(06):1473-1480.
- [8] LI Z, CAI R, FU T Z J, et al. Transferable time-series forecasting under causal conditional shift [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(4):1932-1949.
- [9] 刘江晴.金融科技、影子银行与传统金融部门间的风险溢出研究[D].南昌:南昌大学,2024.
- [10] 杨官学,王家栋.基于 GRU 网络的格兰杰因果网络重构[J].软件导刊,2023,22(10):49-57.
- [11] 刘莉雯.基于格兰杰因果检验分析智慧供应链对我国农产品出口外贸流通结构的影响[J].时代经贸,2024,21(11):97-102.
- [12] Granger, C. W. J. Investigating causal relations by econometric models and cross-spectral methods [J]. Econometrica, 1969, 37(3):424-438.