

Chaotic Financial Series Prediction Study Based on Hybrid Model with Neural Network and Reconstructing Phase Space

Xindi Bai, Xiaowei Deng*

School of Physical and Mathematical Sciences, Nanjing Tech University, Nanjing Jiangsu
Email: baixindi_d2@126.com, *deng.xiaowei@163.com

Received: Dec. 4th, 2019; accepted: Dec. 17th, 2019; published: Dec. 24th, 2019

Abstract

Strange attractor of Shanghai Stock Index series from 2010 to 2019 is reconstructed by the method of reconstructing phase space, then shows that the largest exponent of Lyapunov is positive through numerical simulation, which confirms that the index series is chaotic. The optimum embedding dimensionality is selected by method of false neighbors. On this basis, a hybrid prediction model based on neural network with genetic algorithm and local forecasting with Gaussian Kernel is proposed, which can forecast chaotic series by estimating the trajectory of strange attractor. Compared with RNN, this prediction model has better convergence speed and prediction accuracy.

Keywords

Chaotic Time Series, Reconstructed Phase Space, False Nearest Neighbors, Neural Network

基于相空间重构的混合神经网络对金融混沌序列预测研究

白辛迪, 邓晓卫*

南京工业大学数理科学学院, 江苏 南京
Email: baixindi_d2@126.com, *deng.xiaowei@163.com

收稿日期: 2019年12月4日; 录用日期: 2019年12月17日; 发布日期: 2019年12月24日

*通讯作者。

摘要

应用相空间延迟重构技术重构出2010年至2019年上证指数变动序列的奇异吸引子, 并计算得出其最大Lyapunov指数为正, 从而验证了上证指数序列的混沌特性; 利用虚假邻域法确定恰当的嵌入空间维数应用于遗传神经网络并和高斯核权局部预测构造混合预测算法, 实现对奇异吸引子中运动轨道的估计, 从而达到对该混沌序列的预测。实证表明: 与常用的循环神经网络方法相比, 该技术显著提高了学习效率和预测精度。

关键词

混沌时间序列, 相空间重构, 虚假邻域法, 神经网络

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

混沌理论于1963年由美国气象学家爱德华·诺顿·劳仑兹首次提出。该理论解释了确定系统中可能产生的“随机”现象, 并用模型获得了明确的非周期结果。混沌理论的提出, 极大的扩展了非线性问题的研究方法和思路, 目前已被广泛应用于气象、信号、交通、水利、航空航天等领域[1]-[5]。由于大多金融时间序列的走势具有多种不确定性和非线性性, 因此可以看成是混沌时间序列, 于是利用混沌理论对金融时间序列进行预测成为一种新的预测方法被研究。近年来, 随着机器学习方法的日益成熟, 将机器学习与混沌理论相结合研究金融市场的不确定性问题, 已成为学术界的热点之一, 本文正是基于此展开。

2. 文献综述

对于经济金融数据的非线性动力学特征研究主要从两个方面展开, 一是金融数据是否存在混沌现象。高红兵等[6]不仅证实了我国证券市场运行于具有混沌特征, 而且指出上证指数的分形维大约是2.65, 即决定证券市场运行系统的本质因素最终趋于3个。尹海员、华亦朴[7], 刘超[8]等对中国股市的研究也均表明股票市场存在显著的分形特征。第二个方面是若存在混沌现象如何进行预测? 沈菲[9]、卢山[10]对经济数据的混沌现象作了较全面研究, 并就混沌时间序列的常用预测方法, 如, 相空间预测法, 局域预测法及普通人工神经网络预测法进行了比较研究, 结论是平均误差均在3%以内, 特别遗传神经网络的预测误差更小。顾秋宇等[11]基于金融数据的混沌特征, 提出了一种改进的支持向量机方法, 并以宝钢股份的收盘价为例验证了这种改进方法具有较好的预测精度; 王维等[12]针对CPI数据的非线性特征, 利用混沌BP混合算法构建了CPI预测模型; 李哲敏等[13]对农产品时间序列价格数据建立了动态混沌神经网络模型进行预测, 都取得了较好效果。韩慧婧[14]在相空间重构的基础上利用集合经验模态分解, 降低噪声影响, 实现了效果较好的中长期预测。梁春文[15]利用小波神经网络对重构后的欧盟碳排放权交易价格进行预测, 获得了不错的短期预测效果。

总的来看, 将混沌理论与神经网络方法结合是目前预测方法的主流趋势之一, 现有方法主要将相空间重构作为单一的特征提取手段, 辅助支持向量机或人工神经网络等模型进行预测, 这些方法往往抗噪

能力不强, 并且预测结果缺少长期记忆性。为解决上述问题, 本文通过对序列数据进行合理变换, 提出一种将局部预测法和神经网络方法结合的模型, 在神经网络模型良好的非线性预测能力基础上, 通过基于高斯核权函数的局部预测提升抗噪能力和长期记忆性, 最后对 2010 年至 2019 年上证指数进行实证分析, 对比神经网络和局部预测方法, 检验了本文方法的有效性。

3. 数据来源及处理

本文选取 2010 年 1 月 4 日至 2019 年 6 月 3 日(共 2287 个)上证指数日收盘价, 并选取同时段的 CPI 数据, 构造出消除通货膨胀的股指变动序列:

$$X_i = \ln S_i - \ln CPI_i \quad (1)$$

其中 S_i 是日收盘价, CPI_i 是其对应的居民消费价格指数(以 2010 年 1 月为基期)。将数据如此处理是为了降低原数据的混沌程度[16]。数据来源: 锐思金融数据库。计算软件为 MATLAB。处理后的序列如图 1。

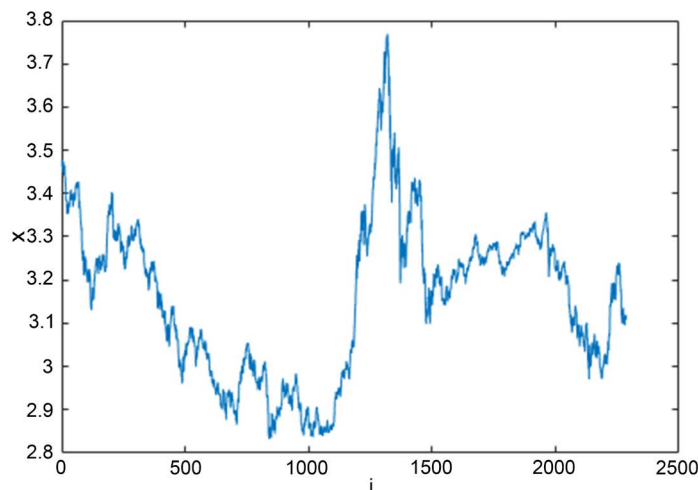


Figure 1. Series of modified Shanghai composite index
图 1. 上证指数变动序列

4. 相空间重构与最大 Lyapunov 指数

4.1. 相空间重构

相空间延迟重构技术的原理在于系统中任意分量的演化是由与其相互作用的其他分量所决定的, 因此这些相关分量的信息就隐含在任意一个分量的发展中。我们可以将一维序列通过时间延迟构造 m 维空间向量 $Y_i = (X_i, X_{i+\tau}, \dots, X_{i+(m-1)\tau})$ 使得这个 m 维相空间与原动力系统拓扑意义下等价, 其中 Takens [17] 证明了在理想情况下只要满足 $m \geq 2d + 1$ (d 是动力系统中吸引子的维数), 就能保证上述相空间的存在性。所以, 针对我们的研究数据, 若具有混沌特征, 只要找到合适的嵌入空间维数 m 与时间延迟 τ 就可以用该序列构造出一个相空间, 而重构出吸引子的相空间在短期预测中具有较高的精度[18]。

当维数 m 不够大时, 奇异吸引子会被“挤压”在低维的相空间中。图 2 给出了 $m = 3$, $\tau = 45$ 时的相图。可以看出相矢基本分两处聚集, 可以判断至少存在两个吸引盘, 并且在两个吸引盘外有一条轨道。恰当的相空间维数 m 和延迟时间 τ 有助于预测混沌状态下金融系统的变化过程。但目前还没有公认确定空间维数 m 与时间延迟 τ 的最优方案, 这些方法要么具有一定的主观性要么会受到噪声的严重影响, 本文采用虚假邻域法确定 m 与 τ 。定义 $W(k, m, \tau)$ 作为在取定 m 与 τ 时与 k 步演化后的距离之比的均值:

$$W(k, m, \tau) = E \left(\sum_{|i-j| \geq k} \ln \frac{\|Y_{i+k} - Y_{j+k}\|}{\|Y_i - Y_j\|} \right) \quad (2)$$

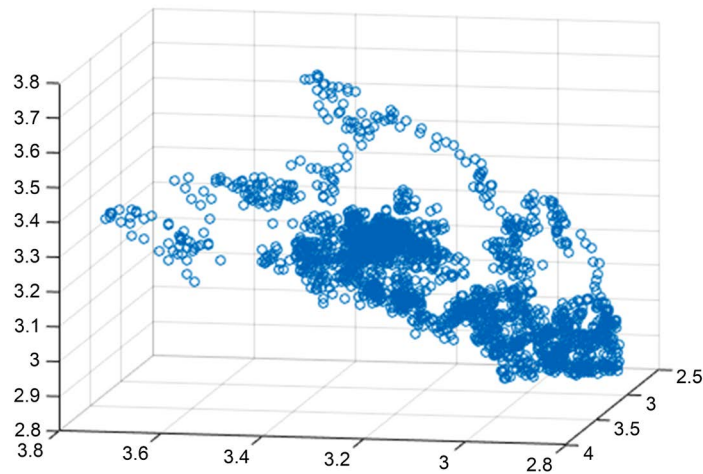


Figure 2. Phase space when $m=3$, $\tau=45$

图 2. $m=3$, $\tau=45$ 对应相空间

一方面, 在维数较低时, 由于奇异吸引子被“挤压”在低维的相空间, 使得原本不相邻的点被投影在同一邻域, 从而形成虚假邻域。但经过(2)的几步迭代后有很大概率使得它们分离, 因此通过计算相空间中各相矢之间 k 步迭代后距离与原距离之比的均值, 当 $W(k, m, \tau)$ 的变化趋于平稳时, 就可确定 m 与 τ 的取值。另一方面, 当嵌入维数 m 足够大时, 奇异吸引子完全展开, W 将不会随着 m 的变化而显著改变, 因此选择使 W 取得最小值的 τ 即可, 特别需要说明的是 k 的取值, k 过小会使得在轨道上运动的相邻相矢误判为虚假邻点, 使得 W 更易受到噪声的影响, 但 k 过大又会影响真实邻域与虚假邻域的区别, 本文中取 $k=4$ 。

通过计算发现当 $1 \leq m \leq 50$ 时, W 均在 $\tau=1$ 时取得最小值, 因此取 $\tau=1$, 计算不同维数下 W 的值, 发现当 $m \geq 10$ 时, W 不再显著变化, 因此确定最佳嵌入维数为 10。结果如图 3。

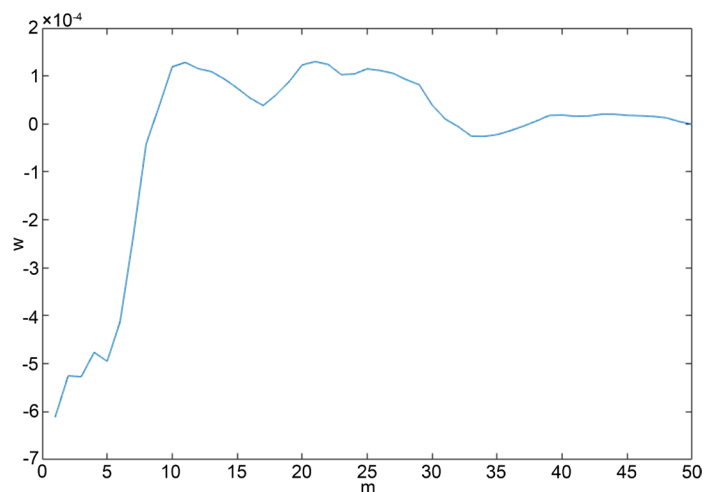


Figure 3. Determining the embedding dimension by flase neighbor

图 3. 虚假邻域法确定嵌入维数

4.2. 最大 Lyapunov 指数

Lyapunov 指数是描述混沌系统特征的一个主要参数, 在重构的相空间中表示两个相矢随时间演化后的发散程度, 即系统对初始条件的敏感程度。负 Lyapunov 指数代表系统受扰动后的恢复时间, 正 Lyapunov 指数代表系统发散速度, 要知道系统是否混沌的, 只需计算最大的 Lyapunov 指数, 大于零即混沌。

通过 Wolf 算法可以计算出最大的 Lyapunov 指数, 首先选取一个相矢及其邻近点, 在演化一定时间间隔后再度量两点间距离, 计算比值, 如果距离超过既定范围则搜索另一个邻点做相同计算, 以此类推穷尽整个序列, 计算出均值即为最大 Lyapunov 指数(LE):

$$LE = \frac{1}{t} \sum_{i=1}^N \ln \frac{l'(t_i)}{l(t_i)} \quad (3)$$

其中 $l(t_i)$ 是 Y_{t_i} 到其最近邻点的距离, $l'(t_i)$ 是距离在既定范围内经过最大迭代次数后两相矢之间的距离。 t 是穷尽整个序列后的总迭代次数。

经计算, 前文所重构的相空间的最大 Lyapunov 指数为 0.0087, 是正数说明本文研究的序列具有混沌特性, 因此, 采用重构相空间对于序列的预测会起到重要作用。

5. 预测算法

5.1. 遗传算法优化的 BP 神经网络

神经网络具有良好的非线性拟合能力, 非常适合对非线性混沌系统的演化进行预测。BP 算法通过训练集学习并将误差反向传递修正神经元之间的权重, 最终达到误差低于要求的结果。但是神经网络的学习结果和效率往往和隐藏层中神经元的拓扑结构相关, 很多隐藏层结构会陷入局部最优从而导致误差无法达到要求, 有时甚至会导致误差随训练而增加。

由上文已经确定混沌系统的最佳嵌入维数为 10, 进而也就确定了神经网络的最佳输入层节点数为 10。神经网络的隐藏层节点数难以确定, 通常采用试凑法, 本文将采用遗传算法进行优化, 避免陷入局部最优, 自动调节至最优结构。算法如下:

- 1) 虚假邻域法确定神经网络输入节点数;
- 2) 随机取初始权值;
- 3) 采用 BP 算法对所有神经网络权值进行学习;
- 4) 选取均方误差小的网络结构直接遗传至下一代;
- 5) 对剩余网络随机进行交叉、变异, 产生子网络作为下一代网络结构, 检验是否达到设定的均方误差要求;
- 6) 重复(2)至(5)直到有神经网络的均方误差达到迭代终止要求。

5.2. 基于高斯核权函数的局部预测算子

局部预测算子利用以当前相矢 Y_N^m 为中心, $r = 3\sigma$ 为半径的邻域内的所有相矢

$$K(u) = \frac{1}{(2\pi)^{m/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} u \Sigma^{-1} u'\right) \quad (4)$$

$$W_{Ni} = \frac{K(Y_N^m - Y_{Ni}^m)}{\sum_{j=1}^n K(Y_N^m - Y_{Nj}^m)} \quad (5)$$

其中 $u \in R^m$ 为 u 的协方差矩阵, Σ 为 u 的协方差矩阵, $K(u)$ 表示迭代距离差为 u 的高斯核函数, 体现邻域内该点在迭代后对 Y_N^m 的反映程度。权值 W_{Ni} 是对邻域内所有点的高斯核函数的归一化。

局部预测结果即为:

$$\hat{Y}_{N+1}^m = \frac{\sum_{i=1}^{n_N} W_{Ni} Y_{N+1,i}^m}{n_N} \quad (6)$$

5.3. 基于局部预测算子和神经网络的混合预测

神经网络算法在非线形拟合上具有优良的效果, 但当输入数据带有噪声干扰时往往会出现难以识别, 误差增大的情况; 而基于高斯核权函数的预测算子在有噪声干扰的情况下可以给出较缓慢的退化[9], 预测结果具有较好的鲁棒性。但是局部预测方法会受到数据混沌特性的影响, 即便只演化一步也有可能产生较大误差。而本文中相空间的最大 Lyapunov 指数为 0.0087, 数值不大, 因此可以考虑将神经网络算法和局部预测方法混合加权预测, 通过神经网络预测降低局部预测中由于混沌性质导致的误差, 同时利用局部预测方法增强预测结果的鲁棒性, 混合预测实施框图由图 4 给出。本文使用训练集数据得出最优权数并将其应用到测试集的预测中, 通过训练数据的计算, 得到局部预测的权重为: $\omega = 0.226$ 。

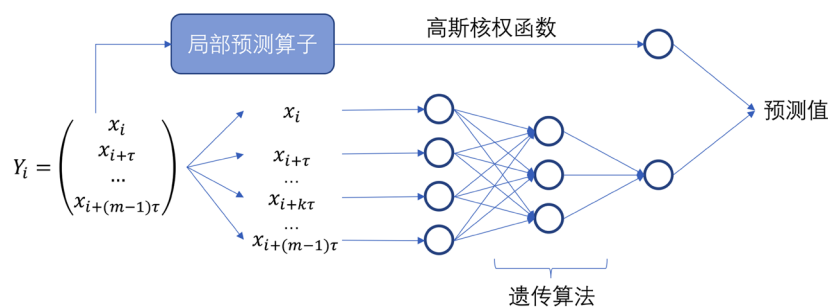


Figure 4. Flowchart of hybrid neural network model

图 4. 混合预测网络结构图

6. 预测

将 2287 个数据重构为 2277 个相空间矢量, 将后 400 个相矢作为测试集数据, 将前 1877 个相矢中随机选取 15% 作为试验集用于判断迭代终止条件。取训练最大步数 3000, 目标均方误差为 10^{-5} 。为了分析比较本文方法的预测结果, 选取循环神经网络(RNN)以及高斯核权函数局部预测结果与本文使用混合预测(Hybrid)进行对比。表 1 给出了三种方法预测结果的各项指标, 图 5 则给出了前两种预测的残差图。

Table 1. Comparison of prediction result

表 1. 各算法预测结果比较

	迭代次数	平均相对误差	最大相对误差	预测涨成功率	预测跌成功率
混合预测	15	1.63%	14.60%	86.36%	93.69%
RNN	47	3.09%	18.42%	72.73%	90.39%
局部预测	-	3.51%	20.47%	84.08%	87.88%

从表 1 可以看出采用遗传算法的 BP 神经网络尽管种群总训练次数很多, 但是利用相空间重构技术选取的最优网络结构仅用 15 次迭代就达到了误差要求, 相比之下 RNN 的收敛速度慢很多。无论从平均

相对误差还是最大相对误差都可以看出,混合预测的精度明显优于 RNN 和单一的高斯核权函数局部预测方法。在预测序列涨的趋势上,成功率显著高于 RNN。

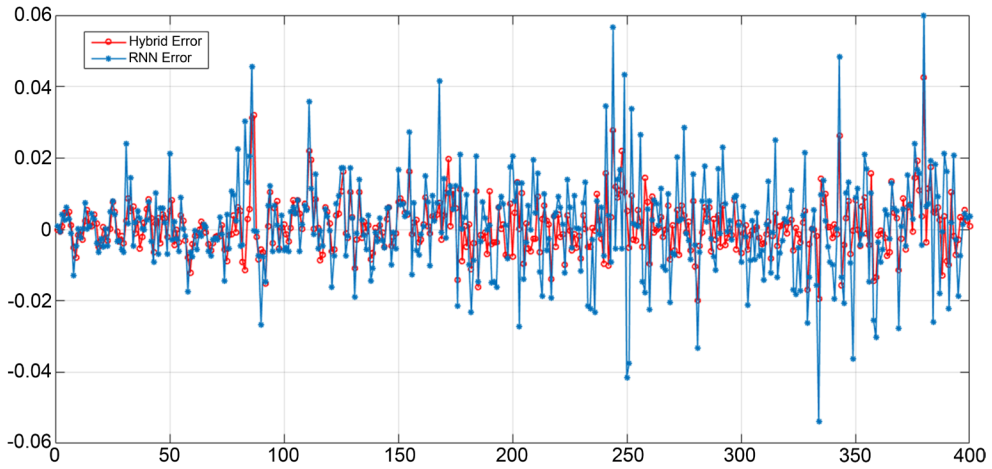


Figure 5. Error comparison of hybrid neural network and recurrent neural network
图 5. 混合神经网络与循环神经网络两种算法预测残差图

从图 5 可以发现在许多区间上本文方法相比循环神经网络能够更好地捕捉到序列变化趋势,虽然两种方法往往会同时低估序列值但是混合预测方法低估幅度更小;但当循环神经网络大幅高估序列值时,混合预测方法的残差基本在 0 附近,给出更为“保守”的预测值,这对于避免金融风险是很有必要的。

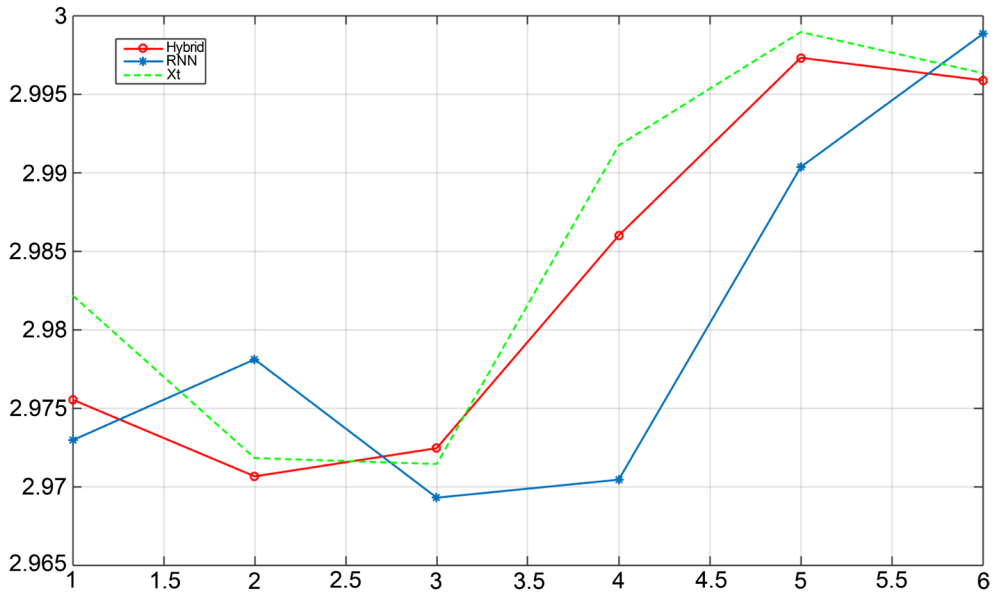


Figure 6. Prediction result comparison of hybrid neural network and recurrent neural network
图 6. 混合神经网络与循环神经网络两种算法预测结果比较

图 6 显示的是混合神经网络与循环神经网络两种模型对未来 6 天序列数据的预测结果,观察这 6 个节点可以发现,混合神经网络在对序列变化趋势的预测上十分准确,而循环神经网络对序列的预测有一定的延迟感,基本符合前一项的变化趋势。从整体上看,混合神经网络不仅有较高的预测精度,而且在数据变化的拐点处具有更好的预测性能。

7. 结论

本文选取 2010 年 1 月 4 日至 2019 年 6 月 3 日上证指数日收盘价和同期定基居民消费价格指数, 构造出一个股指变动序列。通过虚假邻域法, 选取合适的嵌入维数和延迟时间, 重构出该混沌时间序列的相空间并还原了原金融系统的动力学特性, 利用 Wolf 算法计算出最大 Lyapunov 指数为 0.0087, 由此验证了奇异吸引子的存在性。最后提出一种基于遗传算法辅助训练 BP 神经网络和基于高斯核权函数的局部预测算子的混合预测方法, 对相矢演化进行预测并获得序列的预测值, 与以往在时间序列预测中常用的循环神经网络相比, 学习效率和预测精度均有显著提升。

基金项目

江苏省教育厅课题(苏教办外[2017]14 号); 江苏省大学生创新创业训练计划项目(201910291177Y)。

参考文献

- [1] 罗帅, 等. 风速时间序列混沌特征分析及非线性短期预测[J]. 能源工程, 2019(5): 50-56.
- [2] 陈伟, 何家欢, 裴喜平. 基于相空间重构和卷积神经网络的电能质量扰动分类[J]. 电力系统保护与控制, 2018, 46(14): 87-93.
- [3] 李巧茹, 等. 基于相空间重构和 PSO-GPR 的短时交通流预测[J]. 交通信息与安全, 2019, 37(2): 70-76.
- [4] 周建中, 彭甜. 长江上游径流混沌动力特性及其集成预测研究[J]. 长江科学院院报, 2018, 35(10): 1-9.
- [5] 黄飞江, 等. 基于灰色模型和混沌时间序列的卫星钟差预测算法[J]. 电子学报, 2019, 47(7): 1416-1424.
- [6] 高红兵, 潘瑾, 陈宏民. 我国证券市场混沌的判据[J]. 系统工程, 2000, 18(6): 28-32.
- [7] 尹海员, 华亦朴. 我国股票市场流动性的非线性动力学特征研究: 基于分形理论的检验[J]. 管理评论, 2017, 29(8): 43-52.
- [8] 刘超. 证券投资基金系统的非线性实证研究[J]. 管理评论, 2011, 23(12): 18-25.
- [9] 沈菲. 经济系统中的随机分岔与混沌现象研究[D]: [博士学位论文]. 天津: 天津大学, 2005.
- [10] 卢山. 基于非线性动力学的金融时间序列预测技术研究[D]: [博士学位论文]. 南京: 东南大学, 2006.
- [11] 顾秋宇, 朱捷, 徐晶. 混沌时间序列在股票价格长期预测中的应用[J]. 哈尔滨商业大学学报(自然科学版), 2015, 31(1): 108-111+124.
- [12] 王维, 范彦伟. 基于混沌神经网络算法的 CPI 预测[J]. 统计与决策, 2012(10): 22-24.
- [13] 李哲敏, 许世卫, 崔利国, 张建华. 基于动态混沌神经网络的预测研究——以马铃薯时间序列价格为例[J]. 系统工程理论与实践, 2015, 35(8): 2083-2091.
- [14] 韩慧婧. 多尺度下长记忆金融混沌序列预测研究[D]: [硕士学位论文]. 太原: 山西财经大学, 2018.
- [15] 梁春文. 中国与欧盟碳排放权交易价格波动比较研究[D]: [硕士学位论文]. 广州: 暨南大学, 2018.
- [16] Chen, P. (1998) Empirical and Theoretical Evidence of Economic Chaos. *System Dynamics Review*, 4, 81-108. <https://doi.org/10.1002/sdr.4260040106>
- [17] Takens, F. (1981) Dynamic Systems in Turbulence. In: Rand, D. and Young, L.S., Eds., *Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics*, Springer, Berlin, Heidelberg, 366-381. <https://doi.org/10.1007/BFb0091924>
- [18] Procaccia, G.P. (1983) Measuring the Strangeness of Strange Attractors. *Physica D: Nonlinear Phenomena*, 9, 180-208. [https://doi.org/10.1016/0167-2789\(83\)90298-1](https://doi.org/10.1016/0167-2789(83)90298-1)