

基于邻近相点聚类分析的多变量局域多步预测*

宋士豹¹, 杨淑莹²

(1. 天津理工大学 计算机科学与工程学院, 天津 300384; 2. 计算机视觉与系统教育部重点实验室, 天津 300384)

摘要: 针对高维混沌复杂系统的多步预测问题, 提出了一种基于邻近相点聚类分析的多变量局域多步预测模型。首先对于多变量邻近相点的选取, 结合邻近相点多步回溯后的演化规律和变量间的关联信息对演化轨迹的影响, 提出了一种新的多变量演化轨迹相似度综合判据; 然后针对选取全局最优邻近相点耗时长的缺点, 提出了一种基于邻近相点聚类分析的新方案来降低多步预测时间, 提高预测效率。最后通过 Lorenz 混沌数据仿真实验, 表明该模型具有优良的预测性能。

关键词: 聚类分析; 局域模型; 多步预测; 综合判据

中图分类号: TP301.6

文献标志码: A

文章编号: 1001-3695(2018)08-2270-04

doi: 10.3969/j.issn.1001-3695.2018.08.007

Multivariate local multi-step prediction based on cluster analysis of adjacent phase points

Song Shibao¹, Yang Shuying²

(1. School of Computer Science & Engineering, Tianjin University of Technology, Tianjin 300384, China; 2. Key Laboratory of Computer Vision & System for Ministry of Education, Tianjin 300384, China)

Abstract: In order to solve the problem of multi-step prediction for high dimensional chaotic complex systems, this paper proposed a multivariate local multi-step prediction model based on cluster analysis of adjacent phase points. First, it considered the evolution rule of adjacent phase points after multi-step backtracking and the influence of related information on evolutionary trajectory for selecting adjacent phase points. So this paper proposed a new multivariate evolutionary trajectory similarity criterion. Then, the traditional model took a long time to select the optimal phase points. To overcome this shortcoming, it proposed a new method based on cluster analysis of adjacent phase points. This method can reduce the multi-step prediction time and improve prediction efficiency. Finally, the simulation results of Lorenz chaotic data show that this model has good prediction performance.

Key words: cluster analysis; local model; multi-step prediction; synthetic criterion

0 引言

高维混沌时间序列预测一直是非线性预测领域中的研究热点之一^[1], 其在天气预报、径流预测^[2]、经济预测、风电负荷预测^[3]、股票预测^[4]以及目标轨迹预测^[5]等诸多方面均有广泛的应用前景。高维混沌时间序列预测已成为定量研究非线性复杂动力系统的一种有效手段, 其主要预测方法可大致分为全局预测^[6]、自适应预测^[7]、局域预测^[8]三类。全局预测是使用相空间中的全部相点数据来拟合预测函数, 充分利用了所有信息点, 但当数据量较大时, 预测计算量较大、耗时较长、拟合效果欠佳, 并不适合大数据量的高维混沌时间序列预测。自适应预测是使用部分相点数据, 根据当前预测误差来修正拟合函数, 在一步跟踪预测中能达到较高的精度, 但在多步预测中, 预测误差反馈机制存在多步时延, 不能满足拟合函数实时修正的要求。局域预测仅使用邻域中的相点数据来拟合预测函数, 具有所需数据量小、运算速度快、拟合精度高的优点, 其预测性能很大程度上依赖于邻近相点的合理选择, 广泛应用的局域预测模型有局域线性模型^[9]、局域神经网络模型^[10]等。上述缺点均制约着高维混沌时间序列多步预测性能的提高。

近年来, 随着泛函理论的进步, Volterra 滤波器以其训练速度快、非线性逼近能力强、预测精度高等优点得到了广泛的关注^[11]。目前, 国内外学者利用 Volterra 泛函技术对高维混沌时间序列的多步预测进行了一系列的研究, 其应用正在不断深化

与细化。文献[12]提出了局域高阶 Volterra 多步预测, 通过增加 Volterra 滤波器的阶数提高了低维混沌时间序列的多步预测精度, 但系统往往呈现高度的复杂性和非线性, 该方法只利用单变量信息无法完全表征系统的全部动力学特征, 一定程度上制约了多步预测性能的提高, 且对高维混沌时间序列的预测性能有待进一步的验证。文献[13]提出了多变量混沌时间序列 Volterra 多步预测, 该方法将单变量的 Volterra 滤波器扩展为多变量, 考虑了多变量信息对多步预测性能的影响, 在训练样本较少的情况下, 其多步预测性能要优于单变量 Volterra 模型和多变量局域多项式模型, 但由于该模型是基于全局预测法的多步预测, 当数据量较大时, 其多步预测效率有待进一步的改善。文献[14]提出了混沌时间序列遗传算法, 即 Volterra 神经网络 (GA-VNN) 模型, 该方法将 Volterra 泛函模型和 BP 神经网络模型根据两者的结构特点进行了组合, 并用遗传算法来提高组合模型的全局寻优能力, 在风功率超短期多步预测中, 其预测性能明显高于 Volterra 模型和 BP 神经网络模型, 但该组合模型预测步长较短, 应用局限性较大, 结构复杂, 计算量较大, 预测效率较低。以上方法只是从单一角度提高多步预测性能, 并没有从预测精度、预测效率、多变量信息、适用性等综合因素全面考虑。

对高维混沌复杂系统进行局域建模预测是最高效、最常用的方法之一, 而提高局域多步预测精度的关键是选取较合理的邻近相点, 目前主要的判据有欧氏距离^[15]、方向和距离综合判据^[16]、向量相似度^[17]等。欧氏距离仅将相空间中相点的空间

收稿日期: 2017-03-26; **修回日期:** 2017-04-26 **基金项目:** 国家自然科学基金资助项目(61001174); 天津市科技支撑和天津市自然科学基金资助项目(13JCYBJC17700)

作者简介: 宋士豹(1990-), 男, 山东聊城人, 硕士研究生, 主要研究方向为模式识别、时间序列(153127307@stud.tjut.edu.cn); 杨淑莹(1964-), 女, 教授, 主要研究方向为模式识别、时间序列、信号处理。

距离作为选择最优邻近相点的标准,没有考虑相点在相空间中的演化规律。方向和距离综合判据只是将相点的位置信息作为选择标准,但忽略了位置不同的相点对预测点的影响不同。向量相似度只考虑了相点在相空间中一步回溯后的演化规律,但没有考虑变量之间的关联信息对相点演化轨迹的影响。从复杂系统理论可知,多维变量之间的关联信息及其观测数据都参与了系统的演化,以上判据在选取最优邻近相点时,都未全面考虑变量间的关联信息、相点多步回溯后的演化规律、不同位置相点的作用力等综合因素。同时,对于最优邻近相点的选取,传统局域预测法都是将预测相点与相空间中的所有观测相点进行相似度比较,计算量较大。尤其是在数据量较大的观测相点中选取最优邻近相点,运算时间较长,极大地降低了多步预测效率,目前相关文献还没有行之有效的解决方法。

针对以上问题,本文将单变量 Volterra 模型拓展为多变量,并在多变量局域 Volterra 多步预测模型的基础上,结合邻近相点多步回溯后的演化规律和变量间的关联信息对相点演化轨迹的影响,提出了一种新的多变量演化轨迹相似度综合判据,通过选取较合理的邻近相点,来提高混沌时间序列的多步预测精度;依据相空间中观测相点间的相似度特征,先对观测相点进行聚类分析,然后预测相点只与距离最近的簇内相点进行相似度比较,来选取最优邻近相点,有效减小计算迭代次数,降低多步预测时间。利用相点聚类分析的思想,来提高混沌时间序列的多步预测效率,最终实现综合改善多变量局域 Volterra 模型多步预测性能的目的。

1 多变量局域 Volterra 预测模型

1.1 多变量相空间重构

多变量相空间重构技术是多变量复杂混沌时间序列预测的前提。根据 Takens 定理,只要找到一个合适的嵌入维,即 $m \geq 2D + 1$,重构后的相空间将具有与原动力系统相同的几何性质,并与原动力系统在拓扑意义下等价^[18]。设观测到的 M 维混沌时间序列及其重构相空间如下所示:

$$x(i) = (x_1(i), x_2(i), \dots, x_M(i)), i = 1, 2, \dots, N \quad (1)$$

$$V_n = [x_1(n), x_1(n - \tau_1), \dots, x_1(n - (m_1 - 1)\tau_1), \\ x_2(n), x_2(n - \tau_2), \dots, x_2(n - (m_2 - 1)\tau_2), \\ \dots$$

$$x_M(n), x_M(n - \tau_M), \dots, x_M(n - (m_M - 1)\tau_M)]^T \quad (2)$$

其中: τ_i, m_i 分别表示第 i 个变量的延迟时间和嵌入维数, $n = J_0, J_0 + 1, \dots, N; J_0 = \max_{1 \leq i \leq M} (m_i - 1)\tau_i + 1$ 。

1.2 多变量局域 Volterra 预测模型

由于 Volterra 泛函按无穷级数展开很难运用到实际中,将单变量 Volterra 模型拓展为多变量 Volterra 模型,通常采用 m 次求和并二阶截断^[19],其形式如下:

$$\hat{x}(n+1) = Mh_0 + \sum_{i_1=0}^{N-1} h_{11}(i_1)x_1(n-i_1\tau_1) + \\ \sum_{i_1, i_2=0}^{N-1} h_{12}(i_1, i_2)x_1(n-i_1\tau_1)x_1(n-i_2\tau_1) + \dots + \\ \sum_{i_1=0}^{N-1} h_{M1}(i_1)x_M(n-i_1\tau_M) + \sum_{i_1, i_2=0}^{N-1} h_{M2}(i_1, i_2) \\ x_M(n-i_1\tau_M)x_M(n-i_2\tau_M) + e(n) \quad (3)$$

通过状态扩展后,系数总个数为 $h = \sum_{i=1}^M 1 + m_i + m_i(m_i + 1)/2$ 。多变量 Volterra 滤波器的输入矢量为

$$U(n) = [U_1(n), U_2(n), \dots, U_M(n)] \quad (4)$$

$$U_i(n) = [1, x_i(n), x_i(n - \tau_i), \dots, x_i(n - (m_i - 1)\tau_i), \\ x_i^2(n), x_i(n)x_i(n - \tau_i), \dots, x_i^2(n - (m_i - 1)\tau_i)] \quad (5)$$

其中: $U_i(n)$ 表示第 i 个变量的输入矢量, $i = 1, 2, \dots, M$ 。同理,滤波器系数矢量为

$$H(n) = [H_1(n), H_2(n), \dots, H_M(n)] \quad (6)$$

$$H_i(n) = [h_{i0}, h_{i1}(0), h_{i1}(1), \dots, h_{i1}(m_i - 1), \\ h_{i2}(0, 0), h_{i2}(0, 1), \dots, h_{i2}(m_i - 1, m_i - 1)] \quad (7)$$

多变量 Volterra 滤波器输出式(3)可写为

$$\hat{x}(n+1) = H^T(n)U(n) \quad (8)$$

多变量 Volterra 滤波器预测误差为

$$e(n) = x(n+1) - \hat{x}(n+1) \quad (9)$$

2 最优邻近相点的选取

传统邻近相点的选择大多是以混沌轨道一步演化的相似度作为衡量标准,缺少对系统整体演化规律的考虑,忽略了变量间的关联信息对演化轨迹的影响,导致当预测步长较长时,预测轨迹与原轨迹偏离较大。为此,通过定义互信息相对贡献率、多步演化距离相似度和多步演化方向相似度来衡量多变量演化轨迹相似度,提出了一种选择最优邻近相点的综合判据。

互信息能较好地反映变量间的统计依赖程度,可将互信息用于多变量相关性分析^[20]。为了避免直接对高维概率密度函数进行估计,在 K 近邻互信息估计^[21]的基础上,以输入变量对近邻互信息值的相对贡献率来表征变量间的关联信息。 K 近邻互信息估计基本方法为:若以 $\varepsilon(i)/2$ 表示点 $e_i = (x_i, y_i)$ 到它的 K 近邻距离,则映射到 X 轴和 Y 轴的距离分别为 $\varepsilon_x(i)/2, \varepsilon_y(i)/2$, 变量 X 和 Y 的互信息估计值为

$$MI(X, Y) = \psi(k) - \langle \psi(n_x) + \psi(n_y) \rangle + \psi(N) \quad (10)$$

其中: $\psi(x)$ 为 digamma 函数,一般记为 $\psi(x) = \Gamma'(x)/\Gamma(x)$; n_x, n_y 分别为满足条件 $\|x_i - x_j\| \leq \varepsilon_x(i)/2$ 和 $\|y_i - y_j\| \leq \varepsilon_y(i)/2$ 时数据点的个数; $\langle \dots \rangle$ 表示对其中的所有变量取平均。

对于多变量,互信息可由式(10)扩展为

$$MI(X_1, X_2, \dots, X_M) = \psi(k) + (M-1)\psi(N) - \\ \langle \psi(n_{x_1}) + \psi(n_{x_2}) + \dots + \psi(n_{x_M}) \rangle \quad (11)$$

其中: M 为变量的维数。从不同输入变量对整体互信息值的相对贡献率出发,分析多变量间的相关关系。设 $S = \{X_1, X_2, \dots, X_M\}$ 为输入变量子集, S_i 表示在 S 集合中除去 X_i 变量的子集 ($i = 1, 2, \dots, M$), Y 为输出变量。

定义 X_i 变量对互信息值的相对贡献率为

$$V_{\Delta MI}(X_i) = \frac{MI(S, Y) - MI(S_i, Y)}{MI(S, Y)} \times 100\% \quad (12)$$

判据 1 q 步回溯的预测点 $X(p)$ 与相点 $X(t)$ 的距离相似度为

$$\delta(p, t, q) = 1 - \frac{\max(|d_p - d_t|) - |d_p - d_t|}{\max(|d_p - d_t|) - \min(|d_p - d_t|)} \quad (13)$$

其中: $d_p = \|X(p) - X(p - q)\|$, $d_t = \|X(t) - X(t - q)\|$, $\delta(p, t, q) \in [0, 1]$, 且 $\delta(p, t, q)$ 越小,代表邻近相点 $X(t)$ 的 q 步演化距离与 $X(p)$ 的 q 步演化距离越接近。考虑到回溯步长越小的相点,对预测点的作用越大,于是对多步回溯相点的作用力进行加权处理。定义回溯步长距离权重向量为 $\gamma(j)$ ($j = 1, 2, \dots, q$), 满足条件: $\gamma(1) \geq \gamma(2) \geq \dots \geq \gamma(q)$, 取 $\gamma(j) = 2(q - j + 1)/q(q + 1)$ 。

对多步演化距离相似度进行加权处理为

$$\delta(p, t) = \sum_{j=1}^q \gamma(j) \delta(p, t, q) \quad (14)$$

判据 2 q 步回溯的预测点 $X(p)$ 与相点 $X(t)$ 的方向相似度为

$$\cos \theta(p, t, q) = 1 - \left| \frac{[F(p, q), F(t, q)]}{\|F(p, q)\| \cdot \|F(t, q)\|} \right| \quad (15)$$

其中: $F(p, q) = X(p) - X(p - q)$; $F(t, q) = X(t) - X(t - q)$; $[\cdot]$ 代表向量的内积; $\|\cdot\|$ 代表向量的距离范数; $|\cdot|$ 代表绝对值。 $\cos \theta(p, t, q)$ 的值越小,代表邻近相点 $X(t)$ 的 q 步演化方向与 $X(p)$ 的 q 步演化方向越接近。定义回溯步长方向权重向量为 $\varphi(j)$ ($j = 1, 2, \dots, q$), 满足条件: $\varphi(1) \geq \varphi(2) \geq \dots \geq \varphi(q)$, 取 $\varphi(j) = 2(q - j + 2)/q(q + 3)$ 。对多步演化方向相似度进行加权处理为

$$\cos \theta(p, t) = \sum_{j=1}^q \varphi(j) \cos \theta(p, t, q) \quad (16)$$

综合判据 多变量演化轨迹相似度。

对于单变量 X_i 的预测点 $X_i(p)$ 与相点 $X_i(t)$ 的演化轨迹相似度为

$$\lambda_i(p, t) = \mu_1 \delta_i(p, t) + \mu_2 \cos \theta_i(p, t) \quad (17)$$

其中: μ_1, μ_2 分别为距离指标与方向指标的权重值, 且 $\mu_1 + \mu_2 = 1$; $i = 1, 2, \dots, M$ 代表变量的个数; $\lambda_i(p, t)$ 的值越小表示单变量 X_i 的预测点 $X_i(p)$ 与相点 $X_i(t)$ 的演化轨迹相似度越高。结合输入变量对互信息值的相对贡献率, 将单变量演化轨迹相似度拓展到多变量, 对于输入变量集 $S = \{X_1, X_2, \dots, X_M\}$ 的预测点 $S_i(p)$ 与相点 $S_i(t)$ 的演化轨迹相似度为

$$\xi = \sum_{i=1}^M V_{\Delta M}(X_i) \lambda_i(p, t) \quad (18)$$

最后, 根据多变量演化轨迹相似度指标筛选出最优相点数据, 进行模型的训练。

3 邻近相点的聚类分析

对于最优邻近相点的选取, 传统局域预测法都是将预测相点与全局的每个观测相点进行相似度比较, 计算时间较长, 预测效率较低。为此, 在 K-means 聚类算法的基础上, 提出了一种基于邻近相点聚类分析的新方案, 先对观测相点进行聚类分析, 然后预测相点只与距离最近的簇内相点进行相似度比较, 选取最优邻近相点, 来降低预测时间。

由于多变量观测相点的复杂性, 可能会形成大小不一、密度不同的簇, K-means 算法对非均匀分布的数据集聚类效果不是很好, 为了防止簇内相点密度过高而对预测精度有影响, 故对簇内误差之和进行加权处理, 定义聚类的准则函数为

$$J = \sum_{j=1}^K \frac{n_j}{N} \left(\sum_{x \in w_j} \|X - \bar{X}_{w_j}\|^2 \right) \quad (19)$$

其中: K 是聚类中心的个数; N 表示数据集元素总个数; w_j 表示第 j 类; n_j 表示第 j 类内元素的个数; \bar{X}_{w_j} 表示第 j 类的聚类中心矢量; X 表示属于第 j 类的元素矢量。

邻近相点的聚类分析是以预测相点作为演化轨迹相似度的参考标准, 对所有观测相点的演化轨迹相似度进行聚类。由式(17), 定义相点的多变量演化轨迹相似度矢量为

$$\Lambda = [\lambda_1(p, t), \lambda_2(p, t), \dots, \lambda_M(p, t)] \quad (20)$$

由式(18), 计算 l 个观测相点的多变量演化轨迹相似度值, 按升序排序为 $[\xi_1, \xi_2, \dots, \xi_l]$ ($\xi_1 \leq \xi_2 \leq \dots \leq \xi_l$), 并以 $[l/K]$ (取整) 的间隔, 依次从中选取 K 个对应相点的演化轨迹相似度矢量 $\bar{\Lambda}(1), \bar{\Lambda}(2), \dots, \bar{\Lambda}(K)$ 作为初始聚类中心, 尽可能使初始聚类中心分散开, 接近最终的聚类中心, 减小聚类的时间。聚类中心的更新公式为

$$\bar{\Lambda}(j) = \frac{1}{n_j} \sum_{\Lambda \in w_j} \Lambda \quad (21)$$

结合输入变量对互信息值的相对贡献率, 由式(12), 定义多变量互信息贡献率矢量为

$$r = [V_{\Delta M}(X_1), V_{\Delta M}(X_2), \dots, V_{\Delta M}(X_M)] \quad (22)$$

定义一种新的运算符号为

$$A \otimes B = [a_1, a_2, \dots, a_M] \otimes [b_1, b_2, \dots, b_M] = [a_1 b_1, a_2 b_2, \dots, a_M b_M] \quad (23)$$

根据相点的多变量演化轨迹相似度聚类特点, 由式(20) ~ (23), 可将式(19)表示为

$$J = \sum_{j=1}^K \frac{n_j}{N} \left(\sum_{\Lambda \in w_j} \|r \otimes [\Lambda - \bar{\Lambda}(j)]\|^2 \right) \quad (24)$$

对于高维混沌系统的多步预测问题, 大量实践表明, 大部分非线性动力系统可用 Volterra 级数表征, 故可用 Volterra 级数展开式来构造多变量混沌时间序列多步预测模型。因此, 为使多变量邻近相点选取得更合理和更高效, 本文提出了基于邻近相点聚类分析的多变量局域 Volterra 多步预测模型, 具体步骤如下:

a) 对输入变量进行选择和多变量相空间重构。

b) 以预测相点作为演化轨迹相似度的参考标准, 计算所有观测相点的多变量演化轨迹相似度矢量 Λ , 作为待聚类的

数据集, 并确定聚类中心个数 K 。

c) 以预测相点作为演化轨迹相似度的参考标准, 计算所有观测相点的多变量演化轨迹相似度值 ξ , 并按升序排列、等间隔取值, 依次从中选取 K 个对应相点的演化轨迹相似度矢量 $\bar{\Lambda}(1), \bar{\Lambda}(2), \dots, \bar{\Lambda}(K)$ 作为初始聚类中心, 形成 K 个簇。

d) 重新分配相点使准则函数值最大程度地减小, 形成 K 个新簇, 并更新各簇的聚类中心。

e) 循环步骤 d), 直到各簇的聚类中心不再发生变化, 聚类完成。

f) 计算预测相点的演化轨迹相似度矢量 Λ , 并与各聚类中心 $\bar{\Lambda}(j)$ 计算距离, 选取距离最近的簇, 并将预测相点与该簇内所有的观测相点进行演化轨迹相似度值 ξ 的比较, 从中选取 d 个最优邻近相点。

g) 将最优邻近相点用于多变量 Volterra 模型的训练, 求出多变量系数矢量 $H(n)$, 由预测相点生成输入矢量 $U(n)$, 得到一步预测值 $\hat{x}(n+1)$ 。

h) 将预测值加入到观测数据, 依次迭代, 最终实现多步预测功能。

本模型多步预测性能取决于最优邻近相点选取的合理性和高效性。

4 仿真及结果分析

4.1 实验数据

实验采用 Lorenz 模型的 X, Y, Z 三个分量作为输入变量, X 分量作为预测变量对本模型进行验证。其中, $c = 10, b = 8/3, a = 34$, 初始序列值为 $[-1, 0, 1]$, 采样间隔为 0.01, 用四阶 Runge-Kutta 算法求解 Lorenz 模型, 每个分量各获得 9 600 个数据点, 取其后的 1 600 个数据点作为仿真数据, 其中前 1 000 个数据点作为训练数据, 后 600 个数据点作为测试数据, 并对其归一化处理。通过饱和关联维数法和互信息法分别确定输入变量的嵌入维数均为 3, 延迟时间均为 10。综合判据中距离与方向的权重值 μ_1, μ_2 分别为 0.45、0.55, Y, Z 变量对互信息值的相对贡献率分别为 28.7%、22.1%。

4.2 验证综合判据的有效性

为验证本文所提出的多变量演化轨迹相似度综合判据的有效性, 与其他常用判据进行了对比, 以预测均方根误差作为评价准则, 最优邻近相点个数设定为 5, 预测模型使用本文模型, 对比结果如表 1 所示。

表 1 采用不同判据对多步预测性能的影响

判据	1 步预测	5 步预测	10 步预测
欧氏判据	2.73×10^{-3}	3.06×10^{-2}	9.85×10^{-2}
方向和距离综合判据	3.58×10^{-4}	1.62×10^{-3}	4.37×10^{-2}
向量相似度	6.09×10^{-4}	4.98×10^{-3}	6.39×10^{-2}
本文综合判据	7.36×10^{-5}	5.81×10^{-4}	2.26×10^{-3}

由表 1 可知, 在 1 步、5 步、10 步预测中依据欧氏距离来选取最优邻近相点得到的预测误差较大、预测性能较差。依据向量相似度、方向和距离综合判据来选取最优邻近相点得到的预测结果相比欧氏距离有所提升, 但多步预测精度都比本文综合判据少一个数量级。很明显, 本文所提出的综合判据能够选取较合理的邻近相点, 来提高模型多步预测精度。

4.3 聚类中心个数的选取

将 Lorenz 模型产生的 1 000 组训练数据进行多变量相空间重构, 共产生 980 个相点, 把第 980 个相点作为预测相点, 前 979 个相点作为观测相点, 以预测相点作为多变量演化轨迹相似度的参考标准, 计算出所有观测相点的多变量演化轨迹相似度矢量作为待聚类的数据集。前 979 个观测相点的演化轨迹相似度矢量如表 2 所示。

为研究聚类中心个数对预测时间和精度的影响, 以 10 步预测均方根误差作为评价准则, 以 600 步预测时间来统计总耗时, 预测模型使用本文模型, 实验结果如表 3 所示。

表2 观测相点的演化轨迹相似度矢量

X 分量	Y 分量	Z 分量
0.058 2	0.082 9	0.164 4
0.044 3	0.074 3	0.120 4
0.032 9	0.062 1	0.083 7
...
0.001 2	0.001 7	0.002 4

表3 聚类中心个数对预测时间和精度的影响

聚类中心个数	预测误差	总耗时/s
2	0.110 7	20.173
4	6.743×10^{-2}	14.389
6	1.861×10^{-2}	10.541
8	7.485×10^{-3}	9.032
10	2.273×10^{-3}	8.635
12	4.429×10^{-3}	8.502
14	8.602×10^{-3}	8.483
16	1.217×10^{-2}	8.478

由表3可知,随着聚类中心个数的不断增加,10步预测均方根误差先减小后增大,600步预测总耗时一直减小,最后稳定在8s左右。因此,选取10个聚类中心预测误差最小,总耗时相对较少。当聚类中心个数为10时,对979个观测相点的多变量演化轨迹相似度矢量进行聚类,得到XY方向的投影如图1所示,聚类中心和相点分布如表4所示。

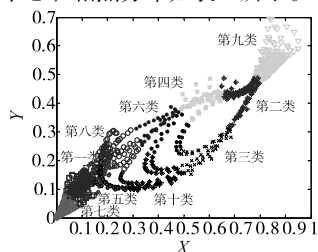


图1 邻近相点聚类的XY方向投影

表4 聚类中心和相点的分布

聚类中心	X 分量	Y 分量	Z 分量	相点数
第一类	0.184 7	0.144 8	0.244 1	149
第二类	0.735 7	0.434 7	0.242 6	108
第三类	0.584 4	0.234 1	0.279 3	51
第四类	0.572 6	0.408 9	0.525 6	47
第五类	0.092 9	0.108 1	0.392 8	133
第六类	0.419 8	0.308 3	0.310 9	57
第七类	0.025 7	0.030 1	0.059 5	150
第八类	0.226 7	0.235 3	0.609 4	81
第九类	0.843 5	0.545 9	0.058 8	149
第十类	0.378 5	0.140 2	0.329 8	54

由于观测相点的多变量演化轨迹相似度矢量的复杂性和不确定性,导致数据集分布不均匀,聚类可能会形成大小不一、密度不同的簇。由图1和表4可知,利用加权处理的K-means算法聚类,避免了簇内元素密度过高,使聚类结果相对均衡,很大程度上减小了对模型多步预测精度的影响。

4.4 各模型多步预测性能比较

为验证本模型的多步预测性能,与其他多步预测模型进行了比较,如图2~5所示。从图2可知,多变量局域线性模型多步预测的有效步长只有120步,120步以后预测轨迹与原轨迹脱离,预测误差波动性较强,误差累积效应严重,预测效果欠佳;从图3可知,本文模型多步预测的有效步长为500步,从200步以后预测轨迹与原轨迹开始偏离,偏离程度先增大后减小,最后恢复正常,并保持较低的预测误差,说明本模型对偏离预测轨迹具有较强的调整能力,很大程度上减弱了误差累积效应,进而增加了预测的有效步长,稳定性较强;从图4和5可知,多变量局域多项式模型和多变量局域神经网络模型多步预测的有效步长分别为290步和220步,虽然在有效步长内预测

误差较小,但同样都存在误差累积效应严重的现象,对偏离预测轨迹的调整能力较弱,限制了预测的有效步长。

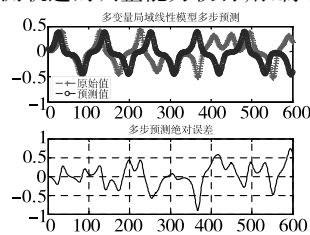


图2 多变量局域线性模型多步预测

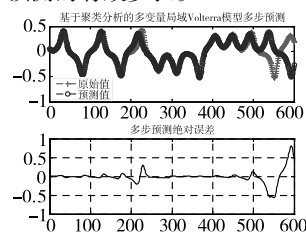


图3 本文模型多步预测

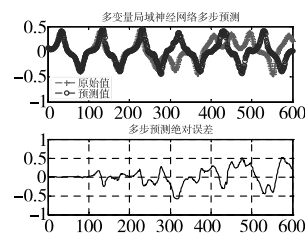


图4 多变量局域神经网络多步预测

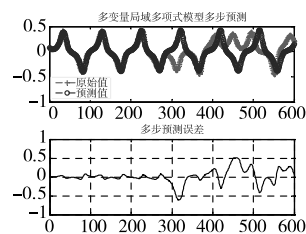


图5 多变量局域多项式模型多步预测

为验证本模型的高效性,以预测均方根误差作为评价准则,以600步预测时间来统计总耗时。各模型预测误差和总耗时比较如表5所示。由表5可知,在多步预测中,本文模型比其他模型预测精度要高1~2个数量级,并在600步的预测中,总耗时最短,只有8.635s,极大地提高了预测效率。综上所述,本文所提出的基于邻近相点聚类分析的多变量局域Volterra多步预测模型具有更优良的多步预测性能。

表5 各模型预测误差和总耗时比较

预测模型	1步预测	50步预测	总耗时/s
多变量局域线性	8.745×10^{-3}	0.1331	9.621
多变量局域神经网络	4.691×10^{-4}	7.634×10^{-2}	50.239
多变量局域多项式	9.957×10^{-4}	8.271×10^{-2}	25.792
本文模型	7.688×10^{-5}	1.843×10^{-3}	8.635

5 结束语

针对高维混沌时间序列特点,本文提出了基于邻近相点聚类分析的多变量局域Volterra多步预测模型,有效地解决了传统局域预测模型选取邻近相点不合理及耗时较长的问题。通过Lorenz混沌数据仿真实验,结果表明:本模型明显提高了多步预测的有效步长和精度,减小了误差累积效应,并大幅降低了多步预测时间,稳定性更强。但同时高维混沌数据中存在较大噪声时,其多步预测性能有待进一步的完善。发展适用性更强、鲁棒性更好、预测性能更优良的混沌时间序列多步预测方法是未来预测领域的热点,也是挑战。

参考文献:

- [1] Findley D F, Ptscher B M, Wei C Z. Modeling of time series arrays by multistep prediction or likelihood methods[J]. *Journal of Econometrics*, 2004, 118(1-2): 151-187.
- [2] Radfar A, Rockaway T D. Captured runoff prediction model by permeable pavements using artificial neural networks[J]. *Journal of Infrastructure Systems*, 2016, 22(3): 04016007.
- [3] Zhou Lin, Li Furong, Tong Xing. Active network management considering wind and load forecasting error[J]. *IEEE Trans on Smart Grid*, 2016, 8(6): 2694-2701.
- [4] Zhang Jian, Li Song. Financial time series analysis model for stock index forecasting[J]. *International Journal of Simulation Systems, Science & Techno*, 2016, 17(16): 12.
- [5] Memon S, Song T L, Kim T H. Smoothing data association for target trajectory estimation in cluttered environments[J]. *EURASIP Journal on Advances in Signal Processing*, 2016, 2016(1): 1-13.
- [6] Horányi A, Cardinali C, Centurioni L. The global numerical weather prediction impact of mean-sea-level pressure observations from drifting buoys[J]. *Quarterly Journal of the Royal Meteorological Society*, 2017, 143(703): 974-985.

(下转第2319页)

表8 各算法在多维多选择背包问题上的平均收敛精度

背包参数设置			Greedy	DPSO	DPSOWSP	QCA
N	M	R				
25	5	5	169.67	163.23	170.32	171.62
50	10	5	351.61	344.71	352.63	355.14
150	15	10	1 430.60	1 604.62	1 620.35	1 624.15
200	20	10	3 374.45	3 525.31	3 482.87	3 501.21
250	25	10	3 577.66	3 759.83	3 876.76	3 890.64
300	30	10	4 200.58	4 680.67	4 726.35	4 750.16
1 000	100	10	2.08E4	2.17E4	2.34E4	2.48E4

横向对比 DPSOWSP、DPSO 和 Greedy 的平均收敛精度, DPSOWSP 在每个测试用例上均优于 DPSO 和 Greedy。该实验结果进一步说明具有多个策略的协同演化算法在各类离散优化问题上比单一策略的演化算法和传统算法更具优势。对比 DPSOWSP 和 QCA 的平均收敛精度, QCA 在维度较低的离散优化问题上收敛优势并不显著;随着问题维度的上升, QCA 的收敛优势越发显著。该实验结果表明采用概率方式协同多个策略比顺序执行的协同方式更适合实际的离散优化问题。

综合上述多维多选择背包问题的实验分析结果,与现有单一策略的演化算法 DPSO、顺序执行的多策略演化算法(DPSOWSP)相比, QCA 更适用于实际离散优化问题的求解,能够在有限时间内获得较高的平均收敛精度。

5 结束语

针对具有时效要求的离散优化问题,本文综合了种群初始化策略、改进的离散粒子群策略,改进的离散单点变异策略和量子策略协同等过程,建立了面向离散优化问题的量子协同演化算法。实验结果显示,已提出的算法能够在有限时间内收敛到精度较高的求解方案,可用于求解具有时效要求的离散优化问题。未来将 QCA 用于求解实际高维离散问题,并通过解空间的分布特征形成启发性信息,促进种群收敛,减缓维度变化对求解效率的消极影响,提高算法的实际应用价值。

参考文献:

- [1] Wong Kachun, Peng Chengbin, Wong M H, *et al.* Generalizing and learning protein-DNA binding sequence representations by an evolutionary algorithm[J]. *Soft Computing*, 2011, 15(8): 1631-1642.
- [2] Wargh S, Kolhe S. Effective semi-supervised approach towards intrusion detection system using machine learning techniques[J]. *International Journal of Electronic Security and Digital Forensics*, 2015, 7(3): 1290-1304.
- [3] Li D, Cheng B, Deng N, *et al.* QoS-aware service composition algorithm and architecture by discrete PSO[J]. *Journal of Computational Information Systems*, 2010, 6(2): 501-511.
- [4] Qin Jin, Lin Xin, Yin Yixin. An algorithmic framework of discrete particle swarm optimization[J]. *Applied Soft Computing*, 2012, 12(3): 1125-1130.
- [5] 崔晓晖, 印桂生, 董红斌. 面向服务匹配问题的协同演化算法[J]. *软件学报*, 2015, 26(7): 1601-1614.
- [6] Dong Hongbin, He Jun, Huang Houkuan, *et al.* Evolutionary programming using a mixed mutation strategy[J]. *Information Science*, 2007, 177(1): 312-327.
- [7] Behnamian J, Fatemi S M T. Development of a PSO-SA hybrid meta-heuristic for a new comprehensive regression model to time-series forecasting[J]. *Expert Systems with Application*, 2010, 37(2): 974-984.
- [8] Idoumghar L, Melkemi M, Schott R, *et al.* Hybrid PSO-SA type algorithms for multimodal function optimization and reducing energy consumption in embedded systems[J]. *Applied Computational Intelligence and Soft Computing*, 2011, 2011(10): 1-12.
- [9] Kathpal S, Vohra R, Singh J, *et al.* Hybrid PSO-SA algorithm for achieving partitioning optimization in various network application [C]// Proc of Conference on Modeling Optimization and Computing. Holland: Elsevier, 2012: 1728-1734.
- [10] Zhang Xin, Dong Hongbin. A new co-evolutionary programming using operators of single-point and particle swarm [C]//Proc of Conference on Modeling Optimization and Computing. Washington DC: IEEE Computer Society, 2010: 5528-5534.
- [11] Bharti K K, Singh P K. Opposition chaotic fitness mutation based adaptive inertia weight BPSO for feature selection in text clustering [J]. *Applied Soft Computing*, 2016, 43(6): 20-34.
- [12] Chen Ning, Gravin N, Lu Pinyan. Competitive analysis via benchmark decomposition [C]// Proc of ACM Conference on Economics and Computation. New York: ACM Press, 2015: 363-367.
- [13] Han Bing, Leblet J, Simon G. Hard multidimensional multiple choice knapsack problems, an empirical study [J]. *Computers & Operations Research*, 2010, 37(1): 172-181.
- [14] 苏爱东, 高维珉, 张永翼. 基于贪心程度和区域界定的预期效率模型求解 0-1 背包问题[J]. *计算机应用研究*, 2015, 32(11): 3304-3308.
- [15] Khan S, Li K F, Manning E G, *et al.* Solving the knapsack problem for adaptive multimedia system[J]. *Studia Informatica, Special Issue on Cutting, Packing and Knapsacking Problems*, 2001, 9: 157-178.
- [16] 程与电子技术, 2009, 31(10): 2476-2479.
- [14] 江岳春, 张丙江, 邢方方, 等. 基于混沌时间序列 GA-VNN 模型的超短期风功率多步预测[J]. *电网技术*, 2015, 39(8): 2160-2166.
- [15] Zhang Xuguang, Ouyang Meiling, Zhang Xufeng. Small scale crowd behavior classification by Euclidean distance variation-weighted network[J]. *Multimedia Tools & Applications*, 2016, 75(19): 11945-11960.
- [16] 杜杰, 曹一家, 刘志坚, 等. 混沌时间序列的局域高阶 Volterra 滤波器多步预测模型[J]. *物理学报*, 2009, 58(9): 5997-6005.
- [17] 王桓, 何怡刚, 侯周国, 等. 基于夹角余弦的混沌局域加权线性预测算法[J]. *高电压技术*, 2009, 35(6): 1483-1487.
- [18] Guo Yina, Liu Qijia, Wang Anhong, *et al.* Optimized phase-space reconstruction for accurate musical-instrument signal classification [J]. *Multimedia Tools & Applications*, 2017, 76(20): 20719-20737.
- [19] Pang Yanjie, Zhang Jiashu. A hierarchical alternative updated adaptive Volterra filter with pipelined architecture[J]. *Digital Signal Processing*, 2016, 56(9): 67-78.
- [20] Pascoal C, Oliveira M R, Pacheco A, *et al.* Theoretical evaluation of feature selection methods based on mutual information [J]. *Neurocomputing*, 2016, 226(2): 168-181.
- [21] Oh B S, Sun Lei, Ahn C S, *et al.* Extreme learning machine based mutual information estimation with application to time-series change-points detection[J]. *Neurocomputing*, 2017, 261(10): 204-206.

(上接第 2273 页)

- [7] Tariq I, Chen Tao, Kirkby N F, *et al.* Modelling and Bayesian adaptive prediction of individual patients' tumour volume change during radiotherapy [J]. *Physics in Medicine & Biology*, 2016, 61(5): 2145-2161.
- [8] Vinoth K C, Natarajan V. Hybrid local prediction error-based difference expansion reversible watermarking for medical images[J]. *Computers & Electrical Engineering*, 2016, 53(7): 333-345.
- [9] Wen Long, Li Xinyu, Gao Liang, *et al.* Surface roughness prediction in end milling by using predicted point oriented local linear estimation method [J]. *International Journal of Advanced Manufacturing Technology*, 2016, 84(9): 2523-2535.
- [10] Patra A, Das S, Mishra S N, *et al.* An adaptive local linear optimized radial basis functional neural network model for financial time series prediction[J]. *Neural Computing & Applications*, 2015, 28(1): 1-10.
- [11] Li Yanling, Zhang Yunpeng, Wang Jing, *et al.* The Volterra adaptive prediction method based on matrix decomposition[J]. *Journal of Interdisciplinary Mathematics*, 2016, 19(2): 363-377.
- [12] Batista E L O, Rui S. A reduced-rank approach for implementing higher-order Volterra filters[J]. *Eurasip Journal on Advances in Signal Processing*, 2016, 2016(12): 118.
- [13] 方芬. 多变量混沌时间序列 Volterra 自适应实时预测[J]. *系统工*