

基于GOSSA和HMM的时间序列预测算法

李大社¹, 孙元威¹, 阮俊虎²

(1. 山东工商学院计算机科学与技术学院, 山东烟台 264000; 2. 西北农林科技大学经济管理学院, 陕西咸阳 712000)

摘要: 时间序列具有非线性和不稳定性等特点, 当前时间序列预测研究面临模型训练参数多、泛化能力差等挑战, 其预测精度无法保证. 基于此, 本文提出一种基于全局最优的麻雀搜索算法(Globally Optimal Sparrow Search Algorithm, GOSSA)和隐马尔可夫模型(Hidden Markov Model, HMM)相融合的时间序列预测模型(GOSSA-HMM). 根据隐马尔可夫模型在模式识别和分类上的优势, 对原始数据做差值处理并划分类别属性, 以此作为隐马尔可夫模型的输入. 采用全局最优的麻雀搜索算法对隐马尔可夫模型的参数进行训练, 以解决参数训练过程中存在的收敛速度慢, 对初始值设置敏感的问题. 将赋予类别属性的差值数据进行分段, 利用改进之后的隐马尔可夫模型测算每段序列走势的概率, 从与当前数据走势相匹配的过去数据集中定位相同的模型实现预测. 通过对山东半岛15个海洋牧场的溶解氧数据进行预测分析, 结果表明与当前主要时间序列预测算法相比, GOSSA-HMM训练的参数较少, 计算成本较低, 具有更好的预测精度和泛化能力.

关键词: 时间序列预测; 隐马尔可夫模型; 麻雀搜索算法

基金项目: 国家重点研发计划项目(No.2020YFD000204); 国家自然科学基金(No.71973106); 山东省重点研发计划(No.2020RKB01555); 烟台市科技创新发展计划项目(No.2021XDHZ062)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2023)09-2492-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220856

Time Series Prediction Algorithm Based on GOSSA and HMM

LI Da-she¹, SUN Yuan-wei¹, RUAN Jun-hu²

(1. School of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong 264000, China;
2. School of Economics and Management, Northwest Agriculture & Forestry University, Xianyang, Shaanxi 712000, China)

Abstract: Time series has the characteristics of non-linearity and instability. There exist some deficiencies in the current researches of time series prediction, such as too many training parameters and poor generalization, which leads to its low prediction accuracy. In order to solve the problems, this paper proposes GOSSA-HMM, a prediction model for time series, based on the fusion of global optimal sparrow search algorithm (GOSSA) and hidden Markov model (HMM). By using the advantages of the hidden Markov model in pattern recognition and classification, the original data in time series can be dealt with by the subtractive preprocessing and classification, which is used as the input of the hidden Markov model. The GOSSA is used to train the parameters of the HMM to solve the problems, such as slow convergence speed and being sensitive to the initial value setting. The D-value data endowed with category attributes are to be segmented, the probability of the trend of each sequence will be calculated by using the improved HMM, and the prediction can be achieved by locating the same pattern with the past datasets matching the trend of the current data. The dissolved oxygen data from 15 marine ranches in the Shandong Peninsula are combined for analysis. Under the same experimental conditions, GOSSA-HMM has fewer training parameters, a lower cost of calculation, and better prediction accuracy and generalization.

Key words: time series; hidden Markov model; sparrow search algorithm

Foundation Item(s): China's National Key R&D Program (No.2020YFD000204); National Natural Science Foundation of China (No.71973106); Key R&D Program of Shandong Province, China (No.2020RKB01555); Yantai Science and Technology Innovation Development Project (No.2021XDHZ062)

1 引言

时间序列数据反映某一个或一些随机变量随时间变化的趋势,而时间序列预测可以从数据中挖掘规律,提供及时和准确的前瞻性信息.时间序列预测方法经历了长时间的发展和完善,已应用于股票和金融市场分析^[1,2]、交通流量预测^[3-5]、电学^[6-8]等领域.

针对时间序列数据,一部分学者采用传统的预测方法,在确定时间序列参数模型的基础上求解模型参数来进行预测. Keshtegar 等^[9]开发了修正响应面法 MRS (Modified Response Surface Method) 和多层感知器神经网络 MLPNN (MultiLayer Perceptron Neural Network) 两种非线性数学建模方法预测溶解氧浓度. Tao 等^[10]开发了一个 C-vine copula 框架来预测河流水温. Yang 等^[11]开发了一种基于特征选择和最小二乘支持向量机的短期电力负荷模型.同时,由于时间序列数据的预测本质上与机器学习方法中的回归分析存在着密不可分的联系,机器学习模型也常用来预测时间序列. Sohrabi 等^[12]用贝叶斯回归方法估计每日河流水温. Kisi 等^[13]开发了基于支持向量机和萤火虫算法的多步提前预测模型用于预测湖泊水位. Olyaie 等^[14]利用三种不同的人工智能方法预测溶解氧浓度.由于时间序列受多种已知和未知因素的共同影响,时间序列往往呈现出稳定性、非线性的特点.通过以上模型进行时间序列建模时会产生较大误差,且泛化能力较差.

随着近年来深度学习网络的发展,多种神经网络模型也被应用在时间序列预测方面.金仓宏等^[15]提出了一种融合时间序列分解策略和时空卷积神经网络的时序预测模型 SDBRNN (Series-Decomposition-Block Recurrent Neural Network). Yu 等^[16]提出了一种基于深度神经网络的股价预测方法. Wan 等^[17]提出了一种 LSTM (Long Short-Term Memory) 新变体 CTS-LSTM (Correlated Time Series Long Short-Term Memory) 用于时间序列预测. Sagheer 等^[18]提出了一种使用深度 LSTM 循环网络的石油产量时间序列预测算法. Jahanbakht 等^[19]提出了一种具有有限元分析的 Transformer 模型预测大堡礁沉积物.谢贵才等^[20]提出一种基于深度学习的多尺度时序卷积网络 MSCNN (Multi-Scale temporal patterns Convolution Neural Networks) 实现人流量时序数据的预测.虽然神经网络在时间序列预测方面能够通过训练从未知数据之间学习到位置关系,进而实现较好的预测精度.但是由于其复杂的网络结构和数量庞大的参数导致其实现预测功能较难,以及实际预测过程中仍然存在一些问题,如模型参数的不可预测性、预测结果的不稳定性和适用性.

隐马尔可夫模型 (Hidden Markov Model, HMM) 是

分析和预测时间序列的一种广泛的模型,由于其训练的参数较少,易于使用和解释,已成功应用于各种类型的时间序列,包括 DNA (Deoxyribo Nucleic Acid) 序列分析^[21-23]、语音信号识别等^[24,25].利用 HMM 进行时间序列预测的方式也有很多,Zhang 等^[26]提出基于高阶 HMM 预测股市价格趋势的方法.对历史中的每日收益进行统计,以证明隐藏状态与价格变化趋势之间的关系. Liu 等^[27]结合 HMM 和高斯混合回归 (Gaussian Mixture Regression, GMR) 进行月流量预测. HMM 使用核化 K-medoids 聚类方法初始化,然后执行 Baum-Welch 算法以学习模型参数. Xiong 等^[28]开发基于高斯混合的 HMM 识别给定分类场景的事故与非事故模式以对事故进行预测.但是 HMM 参数训练过程使用的 Baum-Welch 算法收敛速度慢,对初始值的设置依赖性高,易陷入局部最优解.

由于元启发式算法在优化方面的突出表现,一些研究人员将其应用到 HMM 的参数优化中. Liao 等^[29]构建了一种基于 SA (Simulated Annealing) 和 EM (Expectation-Maximization) 的机械诊断 HMM 启发式优化算法. Tenyakov 等^[30]开发了一个零延迟 HMM 来捕捉外汇汇率数据的演变. Sagayam 等^[31]使用人工蜂群算法优化 HMM 中的状态序列及其模型参数.蚁群算法、人工蜂群算法等元启发式优化算法除了收敛速度慢与陷入局部最优的问题之外,通常还存在种群多样性与收敛速度的矛盾.个体分布越均匀,种群多样性越好,虽得到全局最优解概率大,但是寻优时间变长;个体分布越密集,种群多样性越差,寻优时间虽短,但搜索精度不足.麻雀搜索算法^[32] (Sparrow Search Algorithm, SSA) 是通过模仿麻雀捕食的行为提出的一种新的优化算法,与蚁群算法、人工蜂群算法等元启发式优化算法相比其优势在于局部搜索能力极强,收敛速度较快,但全局搜索与跳出局部最优的能力较弱,因此许多学者对其进行改进来提升寻优性能.李等^[33]提出融合正弦和柯西变异的麻雀搜索算法,虽取得较好效果,但对最优个体位置逃离局部最优的问题缺乏讨论.张等^[34]提出融合多策略的改进麻雀搜索算法,但只考虑迭代前期的聚集情况,缺乏对迭代后期个体逃离局部最优情况的分析.

综上所述,本文提出了一种基于全局最优的麻雀搜索算法 (GOSSA) 和隐马尔可夫模型 (HMM) 的时间序列预测方法 (GOSSA-HMM).首先充分利用 HMM 在模式识别和分类上的优势,对原始数据作差值处理并按照规则赋予其类别属性,可有效减少训练的参数,更好的提取特征信息;其次采用全局最优的麻雀搜索算法训练输入序列的参数,对初始值的设置没有依赖性,加快了算法的收敛速度、避免了算法陷入局部最优解的

缺陷;然后设置时间窗口对数据集进行分段,利用改进之后的HMM计算每段数据趋势出现的概率,在历史区间进行相似程度匹配,解决了时间序列预测易出现滞后性的问题;最后以山东半岛海洋牧场溶解氧数据为例进行测试,验证本算法的可靠性和合理性.

2 隐马尔可夫模型

隐马尔可夫模型(HMM)是一种用来表示双重随机过程的概率模型,如图1所示.假设 Q 是所有可能隐藏状态集合, V 是所有可能观测状态集合,即 $Q=\{q_1, q_2, \dots, q_N\}$, $V=\{v_1, v_2, \dots, v_M\}$.其中, N 是可能的隐藏状态数, M 是可能的观察状态数.对于一个长度为 T 的序列, I 对应状态序列, O 对应观察序列,即 $I=\{i_1, i_2, \dots, i_T\}$, $O=\{o_1, o_2, \dots, o_T\}$,任意一个隐藏状态 $i_t \in Q$,任意一个观察状态 $o_t \in V$.

一个HMM模型可以由隐藏状态初始概率分布 Π ,状态转移概率矩阵 A 和观测状态概率矩阵 B 决定. Π, A 决定状态序列, B 决定观测序列.状态转移概率矩阵 A 为:

$$A=(a_{ij}) \quad (1)$$

其中:

$$a_{ij}=P(i_{t+1}=q_j|i_t=q_i), i=1, 2, \dots, N; j=1, 2, \dots, N \quad (2)$$

表示时刻 t 处于隐藏状态 q_i 的条件下在时刻 $t+1$ 转移到隐藏状态的概率.

观测状态概率矩阵 B 为:

$$B=(b_j(k)) \quad (3)$$

其中:

$$b_j(k)=P(o_t=v_k|i_t=q_j), k=1, 2, \dots, M; j=1, 2, \dots, N \quad (4)$$

表示在时刻 t 处于隐藏状态条件 q_j 下生成观测状态 v_k 的概率.

隐藏状态的初始概率矩阵 Π 为:

$$\Pi=(\pi(i)) \quad (5)$$

其中,

$$\pi(i)=P(i_1=q_i), i=1, 2, \dots, N \quad (6)$$

表示时刻 $t=1$ 处于隐藏状态 q_j 的概率.

因此,HMM模型可以由一个三元组 λ 表示:

$$\lambda=(A, B, \Pi) \quad (7)$$

3 全局最优的麻雀搜索算法

麻雀搜索算法(Sparrow Search Algorithm, SSA)是一种模拟麻雀种群觅食和反捕食行为的元启发式算法.然而麻雀搜索算法存在全局搜索能力弱及易陷入局部最优解的问题,因此本文提出了全局最优的麻雀算法(GOSSA).

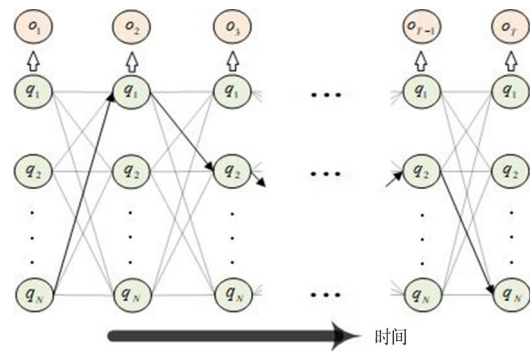


图1 隐马尔可夫模型的基本过程

3.1 引入Sin混沌映射

混沌映射是一种由简单的确定性关系系统产生的随机性序列,被用于生成混沌序列.它常被运用于优化搜索问题,替代伪随机数生成器.本文利用Sin混沌映射替代麻雀搜索算法中随机生成的方式初始化麻雀种群.常见的混沌映射有Logistic映射^[35]、Tent映射^[36]等.Logistic混沌映射的迭代序列是一种两头大中间小的不均匀分布,若其初始值和参数设置不合理,生成的序列容易收敛于某一特定的值.Tent混沌映射具有均匀分布特性,但其满映射范围和参数空间小.两者皆为映射折叠次数有限的模型.Sin一维混沌自映射模型是一种折叠次数无限的模型,杨海东等^[37]通过实验验证,Sin模型拥有比映射折叠次数有限的模型更好的混沌特性.

Sin一维混沌自映射表达式:

$$x_{n+1}=\sin \frac{2}{x_n}, n=0, 1, \dots, N; -1 \leq x_n \leq 1; x_n \neq 0 \quad (8)$$

由于HMM的参数 A, B, Π 的数值范围为 $[0, 1]$,因此本文将Sin混沌映射定义为如下形式,

$$x_{n+1}=\left| \sin \frac{2}{x_n} \right|, n=0, 1, \dots, N; 0 < x_n \leq 1 \quad (9)$$

Sin混沌映射的初始值不能设置为0,防止其在区间内产生不动点和零点.图2表示Sin混沌映射和迭代次数的关系.其中设置初始值为0.555,迭代次数为1000次.图2(a)为连续迭代1000次产生的映射函数值折线图,图2(b)是迭代1000次产生的散点图,由图可知当迭代超过一定次数时,映射函数值将会散布整个解空间,具有良好的均匀分布特性.从Sin混沌映射中生成随机值作为初始化麻雀种群的位置,能提高麻雀搜索算法初始解的质量,提升算法的搜索性能.

3.2 发现者位置更新公式改进策略

麻雀搜索算法中,具有良好适应度值的发现者在搜索过程中会优先获取食物.此外,发现者负责为整个麻雀种群寻找食物并为所有跟随者提供觅食方向.发现者的位置更新公式为:

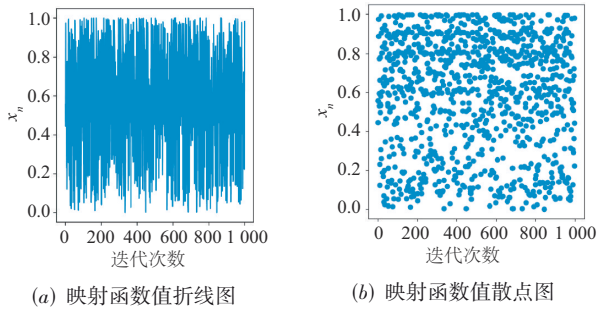


图2 迭代次数与Sin混沌映射关系

$$X_{i,d}^{t+1} = \begin{cases} X_{i,d} \cdot \exp(-\frac{i}{\alpha \cdot \text{Maxiter}}), & R_2 < \text{ST} \\ X_{i,d} + Z \cdot L, & R_2 \geq \text{ST} \end{cases} \quad (10)$$

其中, t 代表当前迭代次数. Maxiter 是一个常数, 表示最大迭代次数. $X_{i,d}^t$ 表示第 t 代中第 i 个个体的第 d 维位置. $\alpha \in (0, 1]$ 是一个随机数. L 为 $1 \times d$ 的全 1 矩阵, 该矩阵内每个元素全部为 1. R_2 ($R_2 \in [0, 1]$) 和 ST ($\text{ST} \in [0.5, 1]$) 分别表示预警值和安全值. Z 是正态分布的随机数. 当 $R_2 < \text{ST}$ 时, 表示当前觅食环境周围没有捕食者, 发现者可以执行更广泛的搜索操作. 当 $R_2 \geq \text{ST}$ 时, 表示种群中的一些麻雀已经发现捕食者, 并向种群其他麻雀发出信号, 此时所有麻雀迅速飞到其他安全的地方进行觅食.

式(10)中, 当 $R_2 < \text{ST}$ 时, 函数 $f(x) = \exp(-\frac{x}{\alpha \cdot \text{Maxiter}})$ 的散点图如图 3(a) 所示, 其中 Maxiter 的值为 1 000, 函数分布随 x 数值的增大, 取值范围由 0 到 1 逐渐缩小至 0 到 0.4 之间. x 值越小, 函数值趋近 1 的概率越大. 随 x 值增大, 函数值分布逐渐均匀. 当 $R_2 < \text{ST}$ 时, 发现者每一维都在缩小, 从迭代开始就向全局最优解靠近, 容易限制搜索范围, 进而陷入局部最优. 本文采用正态分布的函数 $f(x) = 1 + Z$ 优化函数 $f(x) = \exp(-\frac{x}{\alpha \cdot \text{Maxiter}})$, 其中 Z 是正态分布随机数. 其散点图如图 3(b) 所示.

随 x 值增大, 函数值始终在 1 附近的概率最大, 函

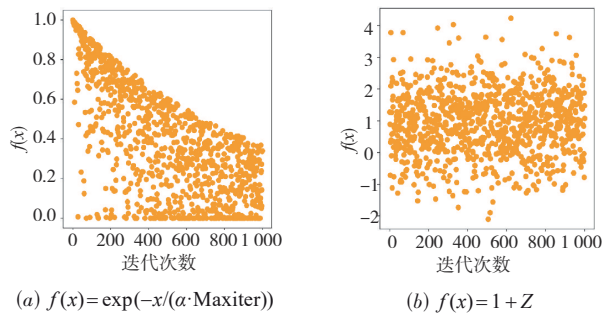


图3 发现者位置更新改进前后对比

数值越远离 1 概率越小. 改进的发现者位置更新公式为:

$$X_{i,d}^{t+1} = \begin{cases} X_{i,d}(1+Z), & R_2 < \text{ST} \\ X_{i,d} + Z \cdot L, & R_2 \geq \text{ST} \end{cases} \quad (11)$$

即去除向原点收敛的过程, 定义为向最优位置移动. 改进后的发现者搜索范围增大, 搜索能力增强, 避免了麻雀个体直接向最优位置跳跃的过程.

3.3 跟随者位置更新公式改进策略

在觅食过程中, 一些跟随者会监视发现者. 一旦察觉到发现者已经找到了更好的事物, 跟随者会立即离开当前位置去争夺食物. 跟随者的位置更新为:

$$X_{i,d}^{t+1} = \begin{cases} Z \cdot \exp(\frac{X_{\text{worst}} - X_{i,d}^t}{i^2}), & i > \frac{n}{2} \\ X_p^{t+1} + |X_{i,d}^t - X_p^{t+1}| \cdot A^+ \cdot L, & i \leq \frac{n}{2} \end{cases} \quad (12)$$

其中, X_p 是当前发现者所在的最优位置, X_{worst} 是当前全局最差位置. A 表示一个 $1 \times d$ 的矩阵, 其中的每个元素随即赋值 1 或 -1, A^+ 满足 $A^+ = A^T (AA^T)^{-1}$. 当 $i > n/2$ 时, 表明适应度值较低的第 i 个加入者争夺不到食物, 此时需要转移到其他地方觅食.

考虑 $|X_{i,d}^t - X_p^{t+1}| \cdot A^+ \cdot L$, 设 $|X_{i,d}^t - X_p^{t+1}| = X = (x_1, x_2, \dots, x_d)$, $A = (a_1, a_2, \dots, a_d)$, 由式 (12) 可知, a_1, a_2, \dots, a_d 的取值为 1 或 -1, 则

$$\begin{aligned} A^+ &= A^T (AA^T)^{-1} \\ &= (a_1, a_2, \dots, a_d)^T ((a_1, a_2, \dots, a_d)(a_1, a_2, \dots, a_d)^T)^{-1} \\ &= (a_1, a_2, \dots, a_d)^T (a_1^2 + a_2^2 + \dots + a_d^2)^{-1} \\ &= \frac{1}{d} (a_1, a_2, \dots, a_d)^T \end{aligned} \quad (13)$$

所以,

$$\begin{aligned} XA^+L &= (x_1, x_2, \dots, x_d) \frac{1}{d} (a_1, a_2, \dots, a_d)^T L \\ &= \frac{1}{d} \sum_{i=1}^d x_i d_i L \end{aligned} \quad (14)$$

其中, $-(\max \{x_i\}) \leq \frac{1}{d} \sum_{i=1}^d x_i d_i \leq \max \{x_i\}$, L 为全 1 矩阵.

由式(12)可知: 当 $i \leq n/2$ 时, 第 t 代麻雀位置与最优发现者位置计算之后的总和被平均分到了最优发现者的位置上作为第 $t+1$ 代麻雀的位置, 且第 $t+1$ 代麻雀位置的每一维都在靠近最优发现者. 同时, 当 $i > n/2$ 时, $0 < \exp(\frac{X_{\text{worst}} - X_{i,d}^t}{i^2}) < 1$, 且越靠近 1 时概率越大. 即该位置的每一维以标准正态分布随机设置在 1 附近.

适应度值适中的跟随者得知发现者找到更好的食物时, 会移动到发现者周围的区域抢夺食物, 使得发现者周围的区域麻雀数量过于密集. 麻雀种群位置缺乏多元化, 呈现显著相似性, 使得算法易陷入局部最优. 适应度值差的麻雀会根据式(10)重置位置. 式(10)一

一定程度上提高了收敛速度,但是没有顾及部分麻雀的搜索区域,搜索精度受到影响.

为克服以上缺陷,本文定义适应度值较低的麻雀为在最优发现者位置随机搜索,定义适应度值适中的麻雀为朝着最优发现者方向搜索.改进的跟随者更新公式为:

$$X_{i,d}^{t+1} = \begin{cases} X_p^{t+1} + N \cdot |X_{i,d}^t - X_p^{t+1}|, & i > \frac{n}{2} \\ X_{i,d}^t + s \cdot |X_{i,d}^t - X_p^{t+1}| \cdot A^+ \cdot L, & i \leq \frac{n}{2} \end{cases} \quad (15)$$

其中, N 是标准正态分布随机数, s 是步长, 值为 0 到 1 之间的随机数. 改进后的跟随者更新公式平衡了所有麻雀的搜索空间, 使得麻雀种群更具多元化, 避免了算法陷入局部最优, 搜索精度进一步提高.

3.4 引入高斯扰动与柯西扰动

由于麻雀搜索算法收敛速度快, 迭代中后期所有麻雀个体在同一位置停留不再转移, 种群呈现强烈的聚集现象而陷入局部最优解, 造成搜索精度不足. 本文引入高斯扰动与柯西扰动策略, 即利用高斯和柯西概率分布的随机数对搜索到的最优解进行干扰.

高斯扰动的公式为:

$$x^{t+1} = x^t \cdot N + x^t \quad (16)$$

x^t 为当前拥有最优适应度值的麻雀位置, x^{t+1} 为高斯扰动之后的麻雀位置, N 是标准正态分布.

柯西扰动的公式为:

$$x^{t+1} = x^t \cdot C + x^t \quad (17)$$

C 为 $\gamma = 1, x_0 = 0$ 的标准柯西分布.

标准正态分布在垂直方向上高于标准柯西分布, 在水平方向上略接近于横轴. 标准柯西概率密度函数在两端上的概率比标准正态分布略大. 根据以上特性, 本文在麻雀算法寻优前中期利用柯西扰动对得到的最优解进行干扰, 扩大搜索范围, 避免算法较早收敛; 而高斯分布在 0 附近的概率偏大, 高斯扰动主要干扰区域在当前最优解的附近, 所以在寻优后期利用高斯扰动进行小范围的搜索以加快收敛速度.

3.5 全局最优的麻雀搜索算法的 HMM 参数估计

HMM 训练算法存在收敛慢、参数随机初始化易陷入局部最优解的问题. 因此, 本文将优化之后的麻雀算法训练 HMM 的参数, GOSSA 具有平衡全局搜索和局部搜索的能力, 有效提高 HMM 的参数训练效果.

将 HMM 的参数估计问题转化成函数优化问题, 再利用 GOSSA 求解函数最优问题. HMM 的参数 A 、 B 、 Π 被转化成麻雀搜索算法中待优化问题的参数, 麻雀搜索算法空间的维数为 A 、 B 、 Π 的元素个数之和, 共 $M \cdot M + M \cdot N + M$ 维. 即 n 只麻雀组成的种群为:

$$X = \begin{bmatrix} a_1^{11} & \cdots & a_1^{MM} & b_1^{11} & \cdots & b_1^{MN} & \pi_1^1 & \cdots & \pi_1^M \\ a_2^{11} & \cdots & a_2^{MM} & b_2^{11} & \cdots & b_2^{MN} & \pi_2^1 & \cdots & \pi_2^M \\ \vdots & & \vdots & & & \vdots & & & \vdots \\ a_n^{11} & \cdots & a_n^{MM} & b_n^{11} & \cdots & b_n^{MN} & \pi_n^1 & \cdots & \pi_n^M \end{bmatrix} \quad (18)$$

X 代表麻雀种群, 每一只麻雀 (矩阵的每一行) 代表搜索空间的一个解, 搜索空间中的最优解确定, 则会确定对应的最优参数 $x = (x^1, x^2, \dots, x^{MM+MN+M})$, HMM 的参数 A 、 B 、 Π 即可确定. 同时本文将 HMM 前向算法得到的 $P(O|\lambda)$ 的值作为麻雀搜索算法中的个体适应度值, 即令目标函数为 $f(x) = P(O|\lambda)$, 其中 O 是已知的一组真实值序列. 通过求解 $f(x)$ 数值最小时对应的麻雀个体的位置, 就能确定对应的 HMM 的参数 $\lambda = \{A, B, \Pi\}$.

GOSSA 的流程如下:

Step1 输入麻雀的数量 n , 发现者比例 PD, 安全值 ST, 侦察者比例 S, 解的维数, 最大迭代次数 Maxiter. 利用 Sin 混沌一维自映射初始化麻雀种群的位置, 生成 n 只麻雀的种群 X , 每只麻雀有 $M \cdot M + M \cdot N + M$ 维.

Step2 根据目标函数计算麻雀搜索算法个体适应度值 f_i , 对适应度值进行排序确定最优适应度值 f_g 和最差适应度值 f_w , 并标记对应的麻雀位置 X_{best} 和 X_{worst} .

Step3 根据改进的发现者位置更新公式更新发现者的位置、根据改进的跟随者位置更新公式更新跟随者的位置、根据侦察者位置更新公式更新侦察者位置.

Step4 若当前迭代次数 t 小于最大迭代次数的一半, 对当前最优位置进行柯西扰动; 若当前迭代次数 t 大于最大迭代次数的一半, 对当前最优位置进行高斯扰动.

Step5 重新确定适应度值并进行排序, 重新确定当前最优适应度值 f_g 和最差适应度值 f_w 及对应的麻雀位置 X_{best} 和 X_{worst} .

Step6 若达到最大迭代次数 Maxiter, 输出最优适应度值对应的麻雀位置作为最优解, 即 HMM 的参数; 否则重复 Step 3、Step 4、Step 5.

4 GOSSA-HMM 预测模型

本文基于 HMM 对于分类识别的优势, 真实值进行差分处理, 将差值作为类别预测下一时刻的差值, 实现真实值的预测. 在此过程中, 改进的 HMM 以一种新的方式参与预测过程.

4.1 数据处理

由于时间序列的真实值受到各种人为或自然因素的影响, 其大多表现为非线性的特点, 所以在差分之前先对数据进行预处理. 并且此时的数据波动较大且不规律, 利用 HMM 直接训练真实值序列, 容易影响模型的预测精度.

$$d_{n-1} = DO_n - DO_{n-1}, n = 2, 3, \dots, N \quad (19)$$

其中, DO_n 表示第 n 个时刻的数据, N 为训练数据集的长度.

4.2 GOSSA-HMM 的预测流程

本文提出的 GOSSA-HMM 模型的预测流程如图 4 所示. 首先对原始数据作差值处理并按照规则赋予其类别属性作为 HMM 的输入. 由于 HMM 作为统计模型, 其在模式分类上具有较强优势, 对原始数据进一步处理有利于减少类别数量, 提升模型识别能力. 其次, 将类别序列

输入 GOSSA 中训练, 将得到的最优种群位置作为 HMM 预测时的参数. 利用 GOSSA 代替传统 HMM 的参数训练方法, 可以有效避免参数初始化的敏感性, 解决算法陷入局部最优解问题, 收敛速度更快. 最后, 利用 HMM 计算每段测试序列的得分, 在训练序列中进行相似程度匹配. 由于利用 GOSSA 得到 HMM 的最优参数, 因此每段序列的得分都是精确的, 能够最大程度寻找到最相似序列, 可解决时间序列预测易出现滞后性的问题.

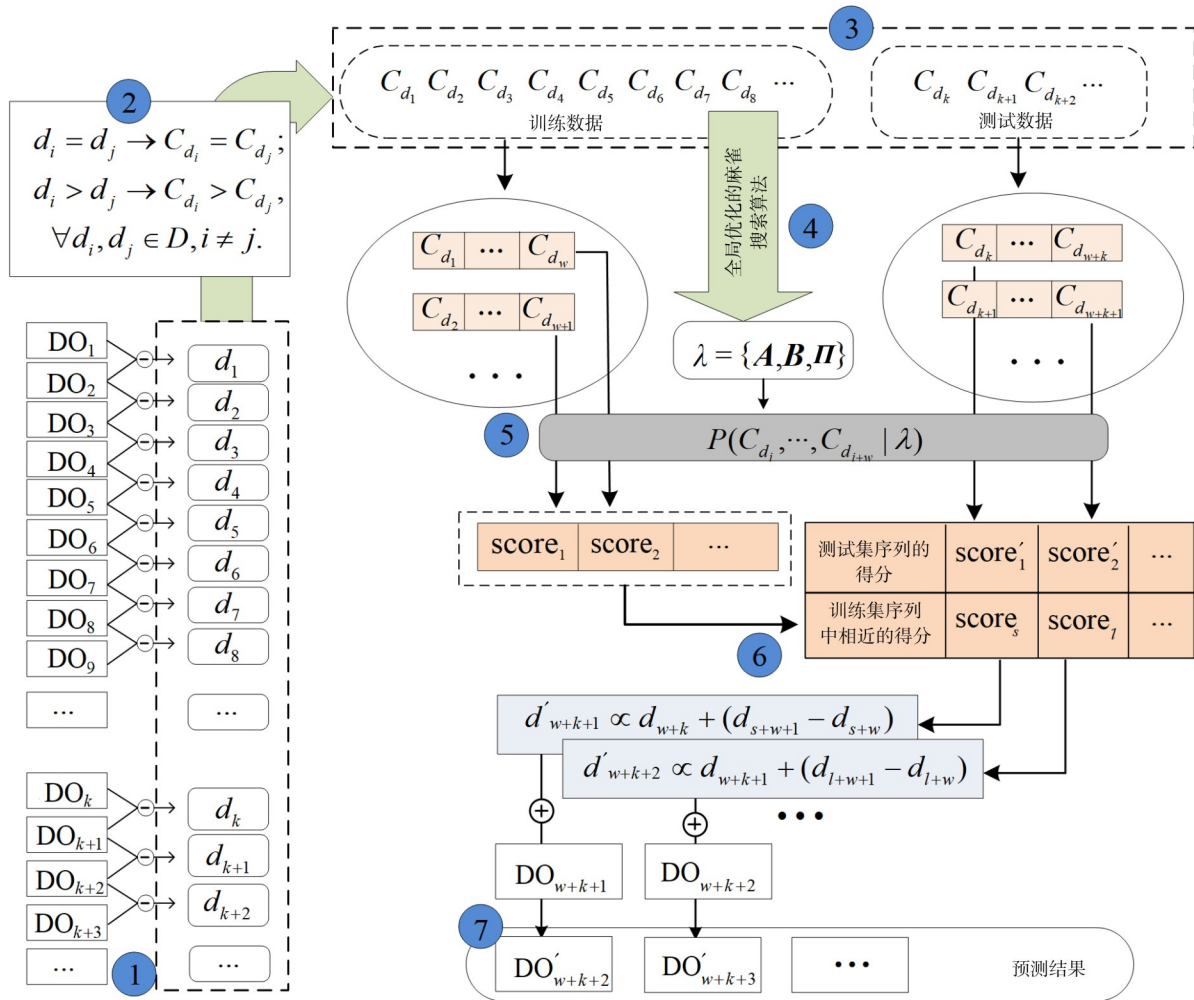


图 4 GOSSA-HMM 预测模型流程

具体流程如下:

Step1 根据式(19)对原始时间序列进行一阶差分.

Step2 对差分之后的数据 $D=\{d_1, d_2, \dots, d_n\}$ 从小到大排序, 并划分到类别 $C=\{1, 2, \dots, m\}$ 中, m 为 D 中不同种类的个数. 划分标准为: $\forall d_i, d_j \in D$, 若 $d_i = d_j$, 则 $C_{d_i} = C_{d_j}$; 若 $d_i < d_j$, 则 $C_{d_i} < C_{d_j}$. 其中 C_{d_i}, C_{d_j} 为 d_i, d_j 的类别.

Step3 通过 Step2 的划分方式, 找到原始差分序列中每个数值对应的类别, 将原始差分序列定义为类别序列.

Step4 训练 HMM 参数. 将转换后的类别序列作

为训练集输入到 GOSSA 中, 按照前文方式估计 HMM 的参数集 λ .

Step5 划分时间窗口, 计算每个时间窗口的似然值. 采用动态训练的方式, 随着时间 t 的移动, 可以获取相应的类别序列 $\{C_{d_i}\}_{t-w+1}^t$. 每组类别序列 $\{C_{d_k}\}_{t-w+1}^t$ 的长度 w 为时间窗口的大小. 将类别序列集合 $\{\{C_{d_k}\}_{t-w+1}^t, t=w, w+1, \dots, n-1\}$ 输入到训练好的 HMM 中, 利用前向算法估计每一个时间窗口出现的概率, 将似然值作为该组序列的得分.

Step6 对于将要预测的差值 d'_{w+k+1} , 假设其之前的一组类别序列为 $\{C_{d_s}, \dots, C_{d_{s+w}}\}$, 利用估计的 HMM 参数计算其概率作为得分, 在过去的数据集中定位与该得分最相近的一组类别序列 $\{C_{d_s}, \dots, C_{d_{s+w}}\}$. 由于当前序列的数据走势应该遵循相同的过去某段序列的数据走势, 因此本文认为将要预测的差值 d'_{w+k+1} 满足:

$$d'_{w+k+1} \propto d_{w+k} + (d_{s+w+1} - d_{s+w}) \quad (20)$$

Step7 若 DO_{w+k+1} 已知, 由式(21)可得预测值 DO'_{w+k+2} :

$$DO'_{w+k+2} = d'_{w+k+1} + DO_{w+k+1} \quad (21)$$

5 实验分析

5.1 参数敏感性分析

为研究预测步长、隐藏状态对模型性能的影响, 本文针对预测步长(4~12)、隐藏状态(2~10)进行了 81 次实验. 图 5 为各个参数敏感性分析热力图, 数值为模型的平均绝对误差 (Mean Absolute Error, MAE) (式(22)). 由图 5 可知, 当预测步长设置为 10、隐藏状态数设置为 4 时的模型 MAE 值最低为 0.086, 模型性能最好. 因此将本文模型的预测步长设置为 10, 隐藏状态数设置为 4.

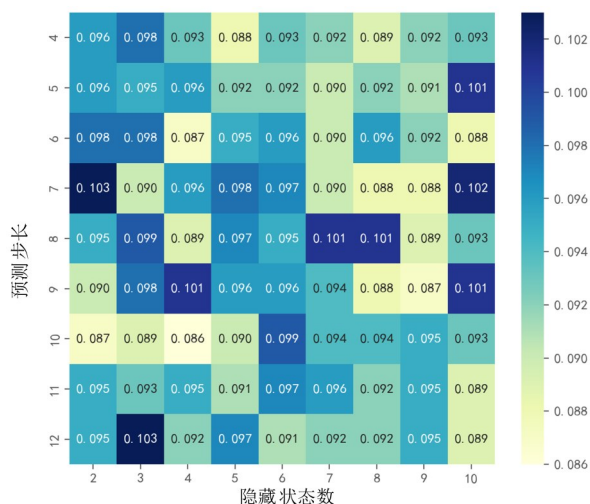


图 5 参数敏感性分析热力图

5.2 单一海洋牧场溶解氧预测性能对比分析

本节以山东半岛某海洋牧场溶解氧数据为研究对象, 采用 Python3.8 语言实现水质参数预测, 在 Windows10 操作系统上运行, 硬件的具体参数为 Intel(R) Core(TM) i9-9900CPU@3.10 GHz. 采用四状态的 HMM 模型, 预测步长设置为 10. 本文使用的数据时间跨度为 2021 年 5 月至 2021 年 10 月, 溶解氧数据每十分钟采集一次, 累计共 25 000 条数据. 在采集数据时, 由于传感器容易受到干扰, 采集的样本数据会存在缺失值和异常值, 为保证模型的预测性能, 利用 Dixon 检验法检测

和剔除异常值, 并利用线性插值法对缺失值进行插补.

此外, 使用文献[34]提出的改进麻雀搜索算法结合 HMM 的预测模型 (ISSA-HMM), 原始 HMM、长短时记忆神经网络 (LSTM)、时间卷积网络 (Temporal Convolutional Network, TCN)、Transformer 五种对比预测模型预测相同的溶解氧数据集. 将 GOSSA-HMM 预测结果与对比模型的预测结果进行比较, 使用平均绝对误差 (MAE)、均方误差 (Mean-Square Error, MSE)、均方根误差 (Root Mean Squared Error, RMSE)、平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 和拟合系数 (R^2) 作为预测评价指标.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{o}_i - o_i| \quad (22)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{o}_i - o_i)^2 \quad (23)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{o}_i - o_i)^2} \quad (24)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{o}_i - o_i}{o_i} \right| \quad (25)$$

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^n (o_i - \hat{o}_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2} \quad (26)$$

其中, o_i 是时刻 i 的真实值, \hat{o}_i 是时刻 i 的预测值, \bar{o} 是真实值的均值. MAE、MSE、RMSE、MAPE 的值越低, R^2 值越高, 模型的性能越好.

表 1 为不同预测模型误差比较, 所有模型均采用十折交叉验证方式, 每个预测模型取十次测试误差的平均值. 结果显示, 在相同条件下, 与传统的 HMM 结果相比, GOSSA-HMM 的 MAE、MSE、RMSE 和 MAPE 指标分别降低了 46.14%、62.5%、35.79% 和 47.27%, 与 ISSA-HMM 结果相比, GOSSA-HMM 的 MAE、MSE、RMSE 和 MAPE 指标分别降低了 10.85%、16.67%、8.69% 和 13.43%, 说明 GOSSA 优化的参数是用于构建时序预测模型更好地选择. 与 LSTM、TCN、Transformer 预测模型相比, GOSSA-HMM 的 MAE、MSE、RMSE 和 MAPE 指标均高于上述三种对比模型, 各个指标都有明显优势. R^2 是常用的评价回归模型优劣程度的指标, R^2 越接近于 1, 表明拟合的效果越好. 表 1 中, GOSSA-HMM 的拟合系数最接近于 1, 表明该模型在溶解氧浓度预测方面具有良好的精度, 能够达到理想的预测效果.

图 6 为不同模型溶解氧浓度预测结果. 可以看出本文所提出的 GOSSA-HMM 模型具有较好的预测能力. 原因是 GOSSA-HMM 引入了全局最优的麻雀搜索算法, 比传统的 HMM 参数训练方法收敛速度更快, 其改进过程解决了算法陷入局部最优解的问题; 且通过匹配当前序列与过去序列, 避免了预测结果的滞后性, 不曾引入无用信

表 1 不同预测模型误差比较

	GOSSA-HMM	ISSA-HMM	HMM	LSTM	TCN	Transformer
MAE	0.041 9	0.047 0	0.077 8	0.114 2	0.071 9	0.094 3
MSE	0.006 0	0.007 2	0.016 0	0.034 9	0.012 5	0.021 7
RMSE	0.077 7	0.085 1	0.126 6	0.187 0	0.112 1	0.147 5
MAPE	0.005 8	0.006 7	0.011 0	0.016 5	0.010 1	0.013 3
R ²	0.995 0	0.992 9	0.986 3	0.969 4	0.988 5	0.981 4

息或丢失重要信息,从而提高了算法的预测精度. ISSA-HMM 预测性能稍差的原因在于虽考虑的麻雀不同种类的更新粗略,但是只考虑迭代前期的种群聚集效应,没有考虑迭代后期的种群聚集效应,从而没有完全解决陷入

局部最优问题. 而 LSTM 模型表现出预测滞后性,这是由于 LSTM 时间序列预测模型在很大程度上取决于逐步预测,随着时间的推移,预测误差累积严重. TCN 在预测溶解氧时,对峰值的预测能力欠缺,TCN 属于卷积神经网络的变种,虽然使用膨胀卷积可以扩大感受野,但是与 LSTM 类似都对数据的要求及需求量很大,其中的池化层会导致部分峰值信息丢失,导致预测效果较差. Transformer 的多头注意力突破了 RNN 模型不能并行计算的限制,但是并不是所有注意力都有用,反而增加了参数数量. Transformer 中还有一个特殊操作是 Sinusoidal 位置编码,利用三角函数生成位置信息与原序列相加虽能识别位置信息但会导致无用信息的增加,所以预测能力稍差.

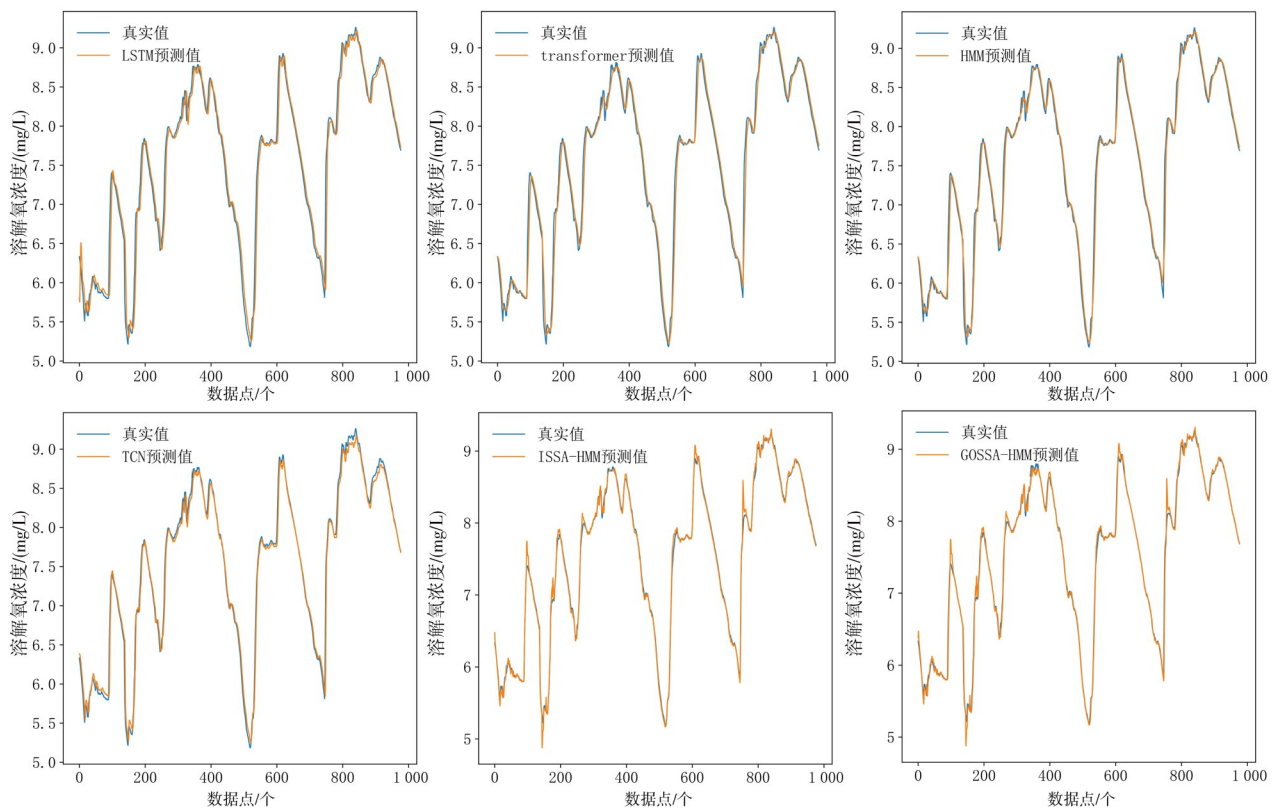


图 6 不同算法模型的真实值与预测值的比较

各个预测模型的复杂度以及时间效率对比如表 2 所示. 复杂度方面, HMM 前向算法的时间复杂度为 $O(TN^2)$, T 为时间序列长度, N 为隐藏层数量. LSTM 包含遗忘门、输入门和输出门. 其中遗忘门操作的时间复杂度为 $O(TN^2)$, 输入门操作的时间复杂度为 $2O(TN^2)$, 输出门操作的时间复杂度为 $O(TN^2)$. TCN 中的主要操作包括膨胀因果卷积、softmax 操作. 膨胀因果卷积操作的时间复杂度为 $O(TN^2)$, softmax 操作的时间复杂度为 $O(T^2)$. Transformer 中的主要操作为自注意力的计算, 而自注意力包括三个步骤: 相似度计算、softmax 操作和加权求和. 相似度计算的时间复杂度为 $O(DT^2)$, D 为

入 Transformer 后嵌入的维度, 一般取 256 或 512. softmax 操作的时间复杂度为 $O(T^2)$. 加权求和的时间复杂度为 $O(DT^2)$. GOSSA-HMM 中也包括前向算法, 时间复杂度为 $O(TN^2)$, GOSSA 的时间复杂度为 $O(nd)$, 其中 n 为种群数量, d 为维度. ISSA-HMM 时间复杂度与 GOSSA-HMM 相同, 都为 $O(TN^2)+O(nd)$. 由于 T, D 的值远大于 N, d, n 的值, 因此, HMM 与 GOSSA-HMM 的模型复杂度最低. 从时间成本上, 也可以看出 HMM 的运行时间最少, 速度最快, 其次是本文提出的 GOSSA-HMM, 两者的运行时间在一个量级上, 说明本文提出的模型训练参数少, 计算成本低.

表 2 不同预测模型复杂度 and 效率对比

	GOSSA-HMM	ISSA-HMM	HMM	LSTM	TCN	Transformer
复杂度	$O(TN^2)+O(nd)$	$O(TN^2)+O(nd)$	$O(TN^2)$	$4*O(TN^2)$	$O(TN^2)+O(T^2)$	$O(DT^2)+O(T^2)+O(DT^2)$
时间成本	147.33 s	153.62 s	96.53 s	7 590.94 s	1 560.08 s	2 365.48 s

5.3 多海洋牧场溶解氧预测性能对比分析

为了验证 GOSSA-HMM 模型的准确性、稳定性以及泛化能力,分别在 15 个不同数据集上进行实验,实验数

据来源于山东半岛各海洋牧场,以 2020 年 5 月至 2020 年 10 月的溶解氧浓度为研究对象. 将 10 月 7 日的真实值与预测值进行比较,实验结果如图 7 所示,横轴代表

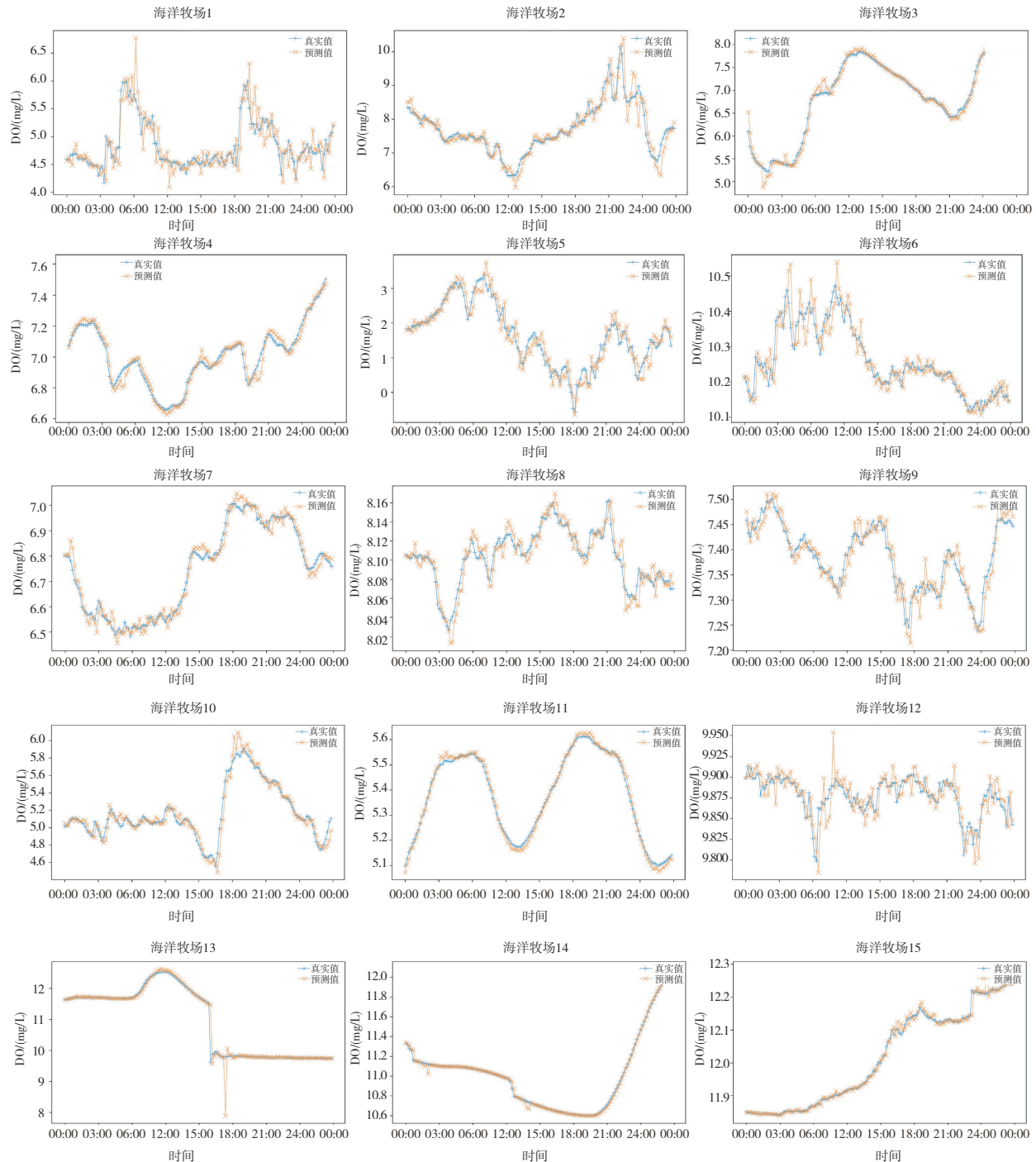


图 7 15 个海洋牧场真实值与预测值比较

时间,纵轴代表溶解氧的浓度.由图 7 可知真实值曲线和预测值曲线几乎重合,偏差较小,表明 GOSSA-HMM 针对不同的训练集和测试集均有较好的预测性能和可靠性.

图 8 为 GOSSA-HMM 在海洋牧场 3 的预测性能泰勒图^[38].泰勒图经常用于评估气候或水文模型,图中的散点代表预测模型,辐射线代表相关系数,横纵坐标代表标准差,弧状虚线代表均方根误差.对于溶解氧浓度的预测,GOSSA-HMM 的标准偏差与真实值的标准偏差最

接近,并且其相关系数最高、均方根误差最小,因此本文提出的模型预测性能最好.而 LSTM 的标准偏差与真实值标准偏差最远,相关系数最低、均方根误差最大,表明 LSTM 预测性能最差.

图 9 为海洋牧场 3 的 MAE、RMSE 和雷达图,用于验证模型的性能.由图可知 GOSSA-HMM 具有最小的 MAE、RMSE 和最大的,表明它是性能最好的模型.相比之下,LSTM 的 MAE、RMSE 最高,值最小,表明它是预测性能最差的模型.

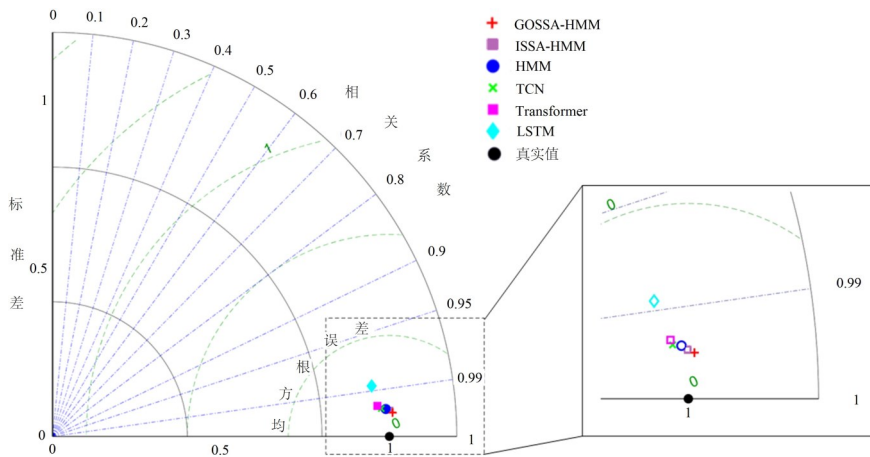


图 8 预测模型性能的泰勒图

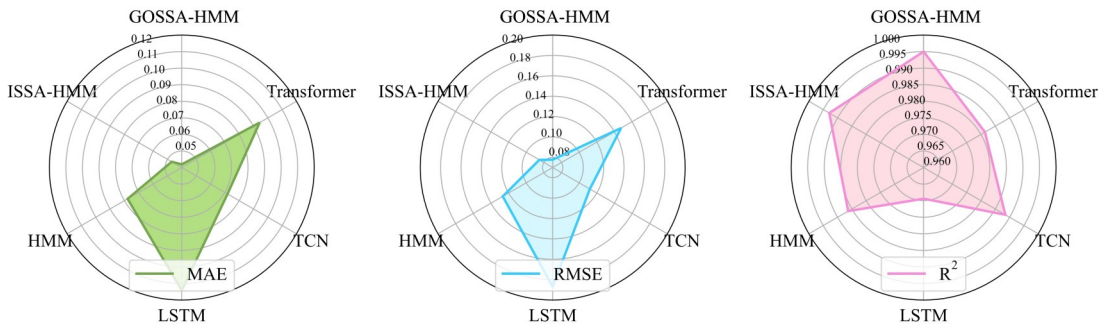


图 9 预测模型性能的雷达图

6 结论

本文构建了一种时序预测模型(GOSSA-HMM),采用全局最优的麻雀搜索算法训练真实值序列的参数,加快了算法的收敛速度,避免了算法易陷入局部最优解的问题;将原始数据进行一阶差分,并赋予其类别属性,充分利用 HMM 在模式匹配方面的优势;采用滑动窗口机制,利用改进的 HMM 极大化每个窗口出现的概率,在历史区间进行相似程度匹配得到预测值,提高了预测精度.将该预测模型应用于山东半岛海洋牧场水体中的溶解氧浓度预测并验证其可靠性.实验结果表明,GOSSA-HMM 模型具有较好的预测精度,可以推广至金融预测、煤炭预测等诸多方

面.然而,为了简化模型,本文提出的 GOSSA-HMM 预测模型没有针对隐藏状态数的设置进行详细分析,并不能保证在所有数据集上呈现最优预测效果.未来,我们将会寻找其他算法来确定给定数据集的最佳 HMM 结构,并探索该方法在更长时间跨度上预测的可能性.

参考文献

[1] TAPIA S, KRISTJANPOLLER W. Framework based on multiplicative error and residual analysis to forecast bitcoin intraday-volatility[J]. Physica A: Statistical Mechanics and Its Applications, 2022, 589: 126613.
 [2] GREENAWAY-MCGREVVY R. Forecast combination for

- VARs in large N and T panels[J]. *International Journal of Forecasting*, 2022, 38(1): 142-164.
- [3] YAN H, ZHANG T A, QI Y, et al. Short-term traffic flow prediction based on a hybrid optimization algorithm[J]. *Applied Mathematical Modelling*, 2022, 102: 385-404.
- [4] 冯宁, 郭晟楠, 宋超, 等. 面向交通流量预测的多组件时空图卷积网络[J]. *软件学报*, 2019, 30(3): 759-769.
FENG N, GUO S N, SONG C, et al. Multi-component spatio-temporal graph convolution networks for traffic flow forecasting[J]. *Journal of Software*, 2019, 30(3): 759-769. (in Chinese)
- [5] 张兴辉, 樊秀梅, 阿喜达, 等. 反向学习的灰狼算法优化及其在交通流预测中的应用[J]. *电子学报*, 2021, 49(5): 879-886.
ZHANG X H, FAN X M, AXIDA, et al. Grey wolf optimization based on opposition learning and its application in traffic flow forecasting[J]. *Acta Electronica Sinica*, 2021, 49(5): 879-886. (in Chinese)
- [6] 蒋锋, 何佳琪, 曾志刚, 等. 基于分解-优化-集成学习方法的电价预测[J]. *中国科学: 信息科学*, 2018, 48(10): 1300-1315.
JIANG F, HE J Q, ZENG Z G, et al. Decomposition-optimization-ensemble learning approach for electricity price forecasting[J]. *Scientia Sinica (Informationis)*, 2018, 48(10): 1300-1315. (in Chinese)
- [7] 郑威迪, 李志刚, 贾涵中, 等. 基于改进型鲸鱼优化算法和最小二乘支持向量机的炼钢终点预测模型研究[J]. *电子学报*, 2019, 47(3): 700-706.
ZHENG W D, LI Z G, JIA H Z, et al. Research on prediction model of steelmaking end point based on LWOA and LSSVM[J]. *Acta Electronica Sinica*, 2019, 47(3): 700-706. (in Chinese)
- [8] 熊有成, 赵鸿. 长短期记忆网络预测混沌时间序列[J]. *中国科学: 物理学 力学 天文学*, 2019, 49(12): 92-99.
XIONG Y C, ZHAO H. Chaotic time series prediction based on long short-term memory neural networks[J]. *Scientia Sinica (Physica, Mechanica & Astronomica)*, 2019, 49(12): 92-99. (in Chinese)
- [9] KESHTEGAR B, HEDDAM S. Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: A comparative study[J]. *Neural Computing and Applications*, 2018, 30(10): 2995-3006.
- [10] TAO Y W, WANG Y K, WANG D, et al. A C-vine copula framework to predict daily water temperature in the Yangtze River[J]. *Journal of Hydrology*, 2021, 598: 126430.
- [11] YANG A L, LI W D, YANG X. Short-term electricity load forecasting based on feature selection and least squares support vector machines[J]. *Knowledge-Based Systems*, 2019, 163: 159-173.
- [12] SOHRABI M M, BENJANKAR R, TONINA D, et al. Estimation of daily stream water temperatures with a Bayesian regression approach[J]. *Hydrological Processes*, 2017, 31(9): 1719-1733.
- [13] KISI O, SHIRI J, KARIMI S, et al. A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm[J]. *Applied Mathematics and Computation*, 2015, 270: 731-743.
- [14] OLYAIE E, ZARE ABYANEH H, DANANDEH MEHR A. A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River[J]. *Geoscience Frontiers*, 2017, 8(3): 517-527.
- [15] 金苍宏, 董腾然, 陈天翼, 等. 融合序列分解与时空卷积的时序预测算法[J]. *电子学报*, 2021, 49(2): 233-238.
JIN C H, DONG T R, CHEN T Y, et al. Spatio-temporal convolutional forecasting based on time-series decomposition strategy[J]. *Acta Electronica Sinica*, 2021, 49(2): 233-238. (in Chinese)
- [16] YU P F, YAN X S. Stock price prediction based on deep neural networks[J]. *Neural Computing and Applications*, 2020, 32(6): 1609-1628.
- [17] WAN H Y, GUO S N, YIN K, et al. CTS-LSTM: LSTM-based neural networks for correlated time series prediction[J]. *Knowledge-Based Systems*, 2020, 191: 105239.
- [18] SAGHEER A, KOTB M. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. *Neurocomputing*, 2019, 323: 203-213.
- [19] JAHANBAKHT M, XIANG W, AZGHADI M R. Sediment prediction in the Great Barrier Reef using vision transformer with finite element analysis[J]. *Neural Networks*, 2022, 152: 311-321.
- [20] 谢贵才, 段磊, 蒋为鹏, 等. 多尺度时序依赖的校园公共区域人流量预测[J]. *软件学报*, 2021, 32(3): 831-844.
XIE G C, DUAN L, JIANG W P, et al. Pedestrian volume prediction for campus public area based on multi-scale temporal dependency[J]. *Journal of Software*, 2021, 32(3): 831-844. (in Chinese)
- [21] TAN C H, LIU X, ZHANG G Y. Inferring brain state dynamics underlying naturalistic stimuli evoked emotion changes with dHA-HMM[J]. *Neuroinformatics*, 2022, 20(3): 737-753.
- [22] GAO R, YU J Y, ZHANG M J, et al. A mathematical formulation of the central dogma of molecular biology[M]// *Nanomedicine*. New York: Jenny Stanford Publishing, 2019: 81-116.

- [23] HAN W, WU Y R, ZENG L T, et al. Building the chordata olfactory receptor database using more than 400, 000 receptors annotated by Genome2OR[J]. *Science China Life Sciences*, 2022, 65(12): 2539-2551.
- [24] 陈哲怀, 郑文露, 游永彬, 等. 标签同步解码算法及其在语音识别中的应用[J]. *计算机学报*, 2019, 42(7): 1511-1523. CHEN Z H, ZHENG W L, YOU Y B, et al. Label synchronous decoding for speech recognition[J]. *Chinese Journal of Computers*, 2019, 42(7): 1511-1523. (in Chinese)
- [25] 龙华, 杨明亮, 邵玉斌. 基于特征流融合的带噪语音检测算法[J]. *通信学报*, 2020, 41(4): 134-142. LONG H, YANG M L, SHAO Y B. Noisy voice detection algorithm based on feature stream fusion[J]. *Journal on Communications*, 2020, 41(4): 134-142. (in Chinese)
- [26] ZHANG M Q, JIANG X, FANG Z H, et al. High-order hidden Markov model for trend prediction in financial time series[J]. *Physica A: Statistical Mechanics and Its Applications*, 2019, 517: 1-12.
- [27] LIU Y Q, YE L, QIN H, et al. Monthly streamflow forecasting based on hidden Markov model and Gaussian mixture regression[J]. *Journal of Hydrology*, 2018, 561: 146-159.
- [28] XIONG X X, CHEN L, LIANG J. A new framework of vehicle collision prediction by combining SVM and HMM [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(3): 699-710.
- [29] LIAO W Z, LI D, CUI S H. A heuristic optimization algorithm for HMM based on SA and EM in machinery diagnosis[J]. *Journal of Intelligent Manufacturing*, 2018, 29(8): 1845-1857.
- [30] TENYAKOV A, MAMON R, DAVISON M. Modelling high-frequency FX rate dynamics: A zero-delay multi-dimensional HMM-based approach[J]. *Knowledge-Based Systems*, 2016, 101: 142-155.
- [31] SAGAYAM K M, HEMANTH D J. ABC algorithm based optimization of 1-D hidden Markov model for hand gesture recognition applications[J]. *Computers in Industry*, 2018, 99: 313-323.
- [32] ZHANG C L, DING S F. A stochastic configuration network based on chaotic sparrow search algorithm[J]. *Knowledge-Based Systems*, 2021, 220: 106924.
- [33] 李爱莲, 全凌翔, 崔桂梅, 等. 融合正余弦和柯西变异的麻雀搜索算法[J]. *计算机工程与应用*, 2022, 58(3): 91-99. LI A L, QUAN L X, CUI G M, et al. Sparrow search algorithm combining sine-cosine and cauchy mutation[J]. *Computer Engineering and Applications*, 2022, 58(3): 91-99. (in Chinese)
- [34] 张晓萌, 张艳珠, 刘禄, 等. 融合多策略的改进麻雀搜索算法[J]. *计算机应用研究*, 2022, 39(4): 1086-1091, 1117. ZHANG X M, ZHANG Y Z, LIU L, et al. Improved sparrow search algorithm fused with multiple strategies[J]. *Application Research of Computers*, 2022, 39(4): 1086-1091, 1117. (in Chinese)
- [35] 王永, 赵毅, Gao Jerry, 等. 基于分段 Logistic 映射的二维耦合映像格子模型的密码学相关特性分析[J]. *电子学报*, 2019, 47(3): 657-663. WANG Y, ZHAO Y, GAO J, et al. Cryptographic feature analysis on 2D coupled map lattices based on piecewise logistic map[J]. *Acta Electronica Sinica*, 2019, 47(3): 657-663. (in Chinese)
- [36] YANG Y, ZHANG H, YAN P, et al. Multi-objective optimization for efficient modeling and improvement of the high temperature PEM fuel cell based Micro-CHP system [J]. *International Journal of Hydrogen Energy*, 2020, 45(11): 6970-6981.
- [37] 杨海东, 鄂加强. 自适应变尺度混沌免疫优化算法及其应用[J]. *控制理论与应用*, 2009, 26(10): 1069-1074. YANG H D, E J Q. An adaptive chaos immune optimization algorithm with mutative scale and its application[J]. *Control Theory & Applications*, 2009, 26(10): 1069-1074. (in Chinese)
- [38] KARGAR K, SAMADIANFARD S, PARSIA J, et al. Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms[J]. *Engineering Applications of Computational Fluid Mechanics*, 2020, 14(1): 311-322.

作者简介



李大社 男, 1978年2月出生于山东临沂, 工学博士, 现为山东工商学院计算机科学与技术学院副教授, 研究方向为智能计算、图像处理等。
E-mail: lidashe@126.com

孙元威 男, 1998年5月出生于山东烟台. 现为山东工商学院计算机科学与技术学院硕士研究生, 主要研究方向为智能计算。
E-mail: 1609456079@qq.com

阮俊虎 男, 1983年10月生于河南周口, 工学博士, 现为西北农林科技大学经济管理学院教授, 研究方向为数字农业、农业大数据与互联网等。
E-mail: rjh@nwsuaf.edu.cn