

基于变分模态分解和引力搜索算法的径流预测模型研究

李晓春¹, 包苑村², 罗军刚², 左岗岗²

(1. 陕西省江河水库工作中心, 陕西 西安 710018;

2. 西安理工大学 西北旱区生态水利国家重点实验室, 陕西 西安 710048)

摘要: 针对月径流数据的数据量小、高度非线性和非平稳性特点,提出了一种基于变分模态分解(VMD)、引力搜索(GSA)与支持向量回归机(SVR)的组合预测模型。首先对月径流样本进行Mann-Kendall(M-K)检验,找出突变点划分训练集与测试集;然后用VMD分别对训练集与测试集进行分解;最后对训练集数据进行训练,利用GSA对SVR的三个主要参数进行寻优,并对测试集的预测结果进行评价。将该模型应用于渭河流域咸阳站与临潼站的月径流预测中。结果表明,VMD-GSA-SVR模型相较于其他模型,有着更高的预测精度与泛化能力,更加符合预测的实际情况,在渭河流域具有较好的可行性与适应性。

关键词: 径流预测; 优化算法; 信号分解; 渭河流域

中图分类号: TV12 文献标志码: A 文章编号: 1006-4710(2021)02-0194-08

Research on runoff prediction model based on variational mode decomposition and gravity search

LI Xiaochun¹, BAO Yuancun², LUO Jungang², ZUO Ganggang²

(1. River and Reservoir Work Center of Shaanxi Province, Xi'an 710018, China; 2. State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi'an University of Technology, Xi'an 710048, China)

Abstract: For the small data volume, highly non-linear and non-stationary characteristics of monthly runoff data, a combined prediction model based on variational mode decomposition (VMD), gravitational search algorithm (GSA) and support vector regression machine (SVR) is proposed. Firstly, Mann-Kendall (M-K) test is carried out on monthly runoff samples to find out break points and divide training set and test set. Then, VMD is used to decompose training set and test set separately. Finally, training set data is trained and GSA is used to optimize the three main parameters of SVR and to evaluate the prediction results of test set. The model is applied to the monthly runoff forecast of Xianyang Station and Lintong Station in Weihe River Basin. The results show that the VMD-GSA-SVR model has higher prediction accuracy and generalization ability than other ones, is more consistent with the actual situation of prediction, and has better feasibility and adaptability in the Weihe River Basin.

Key words: runoff forecast; optimization algorithm; signal decomposition; Weihe River Basin

月径流时间序列受到多种因素的影响和制约,当前流行的数据驱动模型能很好地捕捉径流时间序列的非线性关系,但由于月径流属于小样本,使得数据驱动模型容易陷入过拟合以及局部最小值等情

形^[1]。支持向量回归机(SVR)采用结构风险最小化准则设计学习机器,折衷考虑经验风险和置信范围,在小样本的情况下具有更好的泛化能力^[2]。与此同时,径流时间序列的高度非线性和不稳定性,决定了

收稿日期: 2020-09-23; 网络出版日期: 2020-12-14
网络出版地址: <https://kns.cnki.net/kcms/detail/61.1294.n.20201214.1631.014.html>
基金项目: 国家自然科学基金资助项目(51679186);陕西省自然科学基金基础研究计划资助项目(2019JLZ-15);陕西省水利科技计划资助项目(2018slkj-4);西北旱区生态水利国家重点实验室自主研究课题资助项目(2019KJCXTD-5)
第一作者: 李晓春,男,高级工程师,研究方向为水资源管理与调度。E-mail: 406680783@qq.com
通信作者: 罗军刚,男,博士,教授,博导,研究方向为水资源调度管理和水利信息化。E-mail: jgluo@xaut.edu.cn

需要将 SVR 与其它模型方法相结合,才能更好地提升预测精度。黄巧玲等^[3]将耦合离散小波变换(DWT)与支持向量回归机(SVR)结合,建立了月径流预测的小波支持向量机耦合模型(WSVR),并应用于泾河张家山站的月径流预测中。结果表明,WSVR 模型可以有效提高单一支持向量机模型的预测精度。周有荣等^[4]利用同热传递搜索(SHTS)算法,对混合核支持向量机(SVM)关键参数和混合权重系数进行优化,结果表明,SHTS 算法寻优精度高于 TLBO、GWO 优化算法。李祥蓉^[5]利用静电放电算法(ESDA)优化混合核 SVM 关键参数和混合权重系数,研究结果表明,混合核 ESDA-SVM 模型在预测精度、泛化能力等方面均优于对比模型,具有较好的实际应用价值。王迁等^[6]将粒子群优化算法(PSO)引入到 SVR 模型中,建立了 PSO-SVR 模型,实现了对 SVR 的 RBF 核函数的三个参数的自动优选。实验表明,PSO-SVR 模型较 ANN 模型稳定性更强、可信度更高,且具有更好的应用价值。梁浩等^[7]在优选多元线性回归(MLR)、人工神经网络(ANN)和支持向量机(SVM)单一预报模型的基础上,分别基于经验模态分解(EMD)、集合经验模态分解(EEMD)和小波分解(WD)构建了多种混合模型。结果表明,混合预测模型的预测精度均高于单一模型。

上述研究成果成功将多种优化算法与模型组合,在实例应用中取得了不错的效果。但大多数研究方法在实验开始时就将整个径流时间序列分解,实际带入了径流时间序列的未来信息^[8]。同时,许多优化算法易陷入局部最优^[9]。因此,本文提出基于变分模态分解(VMD)、引力搜索(GSA)与支持向量回归机(SVR)的组合模型。VMD 分别对训练集数据和测试集数据进行分解,GSA 对 SVR 的参数进行全局寻优。将本模型应用于渭河流域临潼站与咸阳站的月径流预测中。实例研究表明,该模型在有效提高预测精度的同时,也更加符合实际预测过程。

1 研究方法

1.1 Mann-Kendall(M-K)检验

在月径流的预测中,通常训练集与测试集的比例为 7:3 左右。由于月径流属于小样本,按比例分配后,测试集的数据量过少,缺乏实际研究意义,因此引入 Mann-Kendall(M-K)检验。

M-K 检验是一种检验水文时间序列的趋势以及突变点的非参数统计检验^[10]。给定月径流时间

序列变量 $x = \{x_1, x_2, \dots, x_n\}$, 则统计量 S_k 为:

$$S_k = \sum_{i=1}^k \sum_{j=i+1}^n a_{ij}, k = 2, 3, \dots, n \quad (1)$$

其中:

$$a_{ij} = \text{sign}(x_j - x_i) = \begin{cases} 1, & x_i < x_j \\ 2, & x_i = x_j \\ 3, & x_i > x_j \end{cases} \quad (2)$$

M-K 检验假设 x 有独立的观测值,如果没有趋势存在,这些观测值分布相同。在此假设下, S_k 的均值和方差分别为:

$$E(S_k) = 0 \quad (3)$$

$$V_0(S_k) = \frac{k(k-1)(2k+5)}{18} \quad (4)$$

x 的检验统计量定义为:

$$UF_k = \begin{cases} S_k - 1/\sqrt{V_0(S_k)}, & S > 0 \\ S_k + 1/\sqrt{V_0(S_k)}, & S < 0 \end{cases} \quad (5)$$

在显著性水平 $1-\alpha(Z_{1-\alpha})$ 下,通过比较 UF_n 与标准正态变量来检测 x 的趋势, $\alpha(0 < \alpha < 0.5)$ 是 M-K 检验错误拒绝原假设 H_0 的容忍概率。当 $|UF_n| \geq Z_{1-\alpha}$ 时,拒绝 H_0 ; 否则,接受 H_0 。在拒绝 H_0 的条件下,如果 $|UF_n|$ 大于(小于)零, x 随时间呈单调增加(减少)的趋势。

为了检测月径流的突变点,将月径流时间序列按相反的顺序进行排序,即 $x = \{x_n, x_{n-1}, \dots, x_1\}$, 重复上述过程以获得测试统计序列 $\{UB_k\}_{k=2}^n$ 。如果序列 $\{UF_k\}_{k=2}^n$ 与 $\{UB_k\}_{k=2}^n$ 在置信区间 $[-Z_{1-\alpha}, Z_{1-\alpha}]$ 内有交点,这些交点就是突变点。在此基础上,将突变点前的月径流数据作为训练集,将突变点后的月径流数据作为测试集,这样既可以避免测试集的数据量过少,也可更好地测试模型的泛化能力。

1.2 变分模态分解(VMD)

鉴于月径流时间序列的高度不稳定性,引入电力系统的信号分解模式^[11],将月径流序列分解为多个分量,以便模型更好地学习月径流的变化规律。

VMD 是一种新的信号分解模式,对原始信号采用非递归和变分模式分解的方法^[12],将输入信号分解成若干个子模式,即周期性的分量(IMF)和一个残差(R)。VMD 可以手动设置分解的模态数 K ,对噪声具有较好的鲁棒性,能显著降低计算的复杂度。因此,VMD 可以对训练集和测试集分别进行分解,以避免模型在训练过程中混入测试集的信息,同时也能保证训练集与测试集的特征维度一致。

VMD 的核心是一个受限的变分问题,将非平稳信号 f 分解为 K 个有限带宽的模态分量,为了保证每个模态分量的估计带宽之和最小,须使所有模态

之和与原始信号相等。Dragomiretskiy 与 Zosso 在 2014 年提出了该受限变分问题^[13]：

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (6)$$

s. t. $\sum_{k=1}^K u_k = f$

式中： u_k 是模态函数的集合； ω_k 是第 k 个模态的中心频率； K 是模态数； $\delta(t)$ 是狄拉克分布； \otimes 是卷积运算； $u_k(t)$ 是各子序列的模态函数； $e^{-j\omega_k t}$ 是复平面上模态函数中心频率的向量描述。

1.3 支持向量回归机(SVR)原理及其参数

支持向量回归机(SVR)首先通过核函数^[14]将低维的非线性回归问题映射至高维的空间,在高维的空间计算回归函数,基于结构风险最小化原则^[15],有效避免因数据量不足而引起的预测精度过低、泛化能力差等问题。具体的计算过程为：

给定训练样本数据 $S = \{(x_i, y_i) | x_i \in \mathbf{R}, y_i \in \mathbf{R}, i = 1, 2, \dots, m\}$, 其中 x_i 为模型输入, y_i 为模型输出, m 为样本数量。则 SVR 的回归函数为：

$$y = w^T \cdot \phi(x) + b \quad (7)$$

式中： w^T 为权重向量； b 为偏置向量； $\phi(x)$ 用于将输入向量映射至高维空间。

假设我们能容忍 $f(x)$ 与 y_i 之间最多有 ϵ 的偏差,即仅当 $f(x)$ 与 y_i 之间差值的绝对值大于 ϵ 时才计算损失,于是 SVR 问题可以转化为：

$$\min_{w, b, \xi_i, \hat{\xi}_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i) \quad (8)$$

其中：

$$f(x) - y_i \leq \epsilon + \xi_i$$

$$y_i - f(x) \leq \epsilon + \hat{\xi}_i$$

$$\xi_i \geq 0, \hat{\xi}_i \geq 0, i = 1, 2, \dots, N, m \leq N$$

式中： C 为惩罚系数,用来保证模型的泛化能力； ϵ 为不敏感损失函数； ξ_i 、 $\hat{\xi}_i$ 为松弛变量,用来保证模型通过核函数升维后的容错性。

运用拉格朗日乘子求解式(8)得：

$$w = \sum_{i=1}^N (\lambda_i - \hat{\lambda}_i) x_i \quad (9)$$

式中： λ_i 、 $\hat{\lambda}_i$ 为拉格朗日乘子。

由 Karush-Kuhn-Tucker^[16] 条件解得的 SVR 模型可以表示为：

$$f(x) = \sum_{i=1}^N (\lambda_i - \hat{\lambda}_i) K(x_i, x) + b \quad (10)$$

式中： $K(x_i, x) = \phi(x_i)^T \phi(x_j)$ 为核函数。

支持向量回归机中常用的核函数包括线性核函

数、多项式核函数和高斯径向基核函数^[17] (RBF)。针对径流预测问题,RBF 核函数具有超参数少、映射维度高与决策边界多样的优点,可以更好地适应非线性的径流预测问题。

RBF 核函数的主要超参数有惩罚系数 C 、不敏感损失函数 ϵ 。除此之外,RBF 核函数还有一个独有参数 $gamma$,隐含地决定了数据映射到新的特征空间后的分布, $gamma$ 值越大,支持向量越少, $gamma$ 值越小,支持向量越多,支持向量的个数会影响训练与预测的速度。

1.4 引力搜索算法(GSA)

引力搜索算法(GSA)是一种新的启发式搜索算法^[18],它受万有引力定律的启发,适用于模型的参数寻优,具有较好的全局搜索能力。GSA 可以描述为一个 n 维的空间中有 s 个粒子,则第 i 个粒子的位置定义为：

$$X_i = (x_i^1, \dots, x_i^d, \dots, x_i^n) \quad (11)$$

式中： x_i^d 表示第 i 个粒子在第 d 维的位置； n 为搜索空间的维度。

在开始搜索前,所有粒子的位置都是随机的。在某一时刻 t , 粒子 i 与 j 之间的引力为：

$$F_{ij}^d(t) = G(t) \frac{M_{pi}(t) M_{ai}(t)}{R_{ij}(t) + \epsilon} [x_j^d(t) - x_i^d(t)] \quad (12)$$

式中： $M_{pi}(t)$ 为施力物体 j 的惯性质量； $M_{ai}(t)$ 为受力物体 i 的惯性质量； $G(t)$ 为引力常数,其值随时间的变化而变化,如式(13)所示(通常设置 G_0 为 100, α 为 10)； $R_{ij}(t)$ 为 i 与 j 之间的欧式距离,如式(14)所示； ϵ 为一个小的常数,用来防止 i 与 j 之间的欧式距离为 0。

$$G(t) = G_0 e^{-\alpha \frac{t}{T}} \quad (13)$$

$$R_{ij}(t) = \sqrt{\sum_{d=1}^n (x_j^d(t) - x_i^d(t))^2} \quad (14)$$

粒子 i 在 t 时刻受到的其它粒子的引力为：

$$F_i^d(t) = \sum_{j=1, j \neq i}^n rand_j F_{ij}^d(t) \quad (15)$$

式中： $rand_j$ 为 $[0, 1]$ 范围内的随机数。

在第 d 维,粒子 i 的加速度为：

$$a_i^d(t) = \frac{F_i^d(t)}{M_{ii}(t)} \quad (16)$$

式中： $M_{ii}(t)$ 为粒子 i 的惯性质量。

假设重力质量和惯性质量相等,则重力质量和惯性质量可更新为：

$$M_{ai} = M_{pi} = M_{ii} = M_i, i = 1, 2, \dots, n \quad (17)$$

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \quad (18)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^n m_j(t)} \quad (19)$$

式中: $fit_i(t)$ 为粒子 i 在 t 时刻的适应度值。
 $worst(t)$ 与 $best(t)$ 在求解最小值问题时定义为:

$$best(t) = \min_{j \in \{1, \dots, n\}} fit_j(t) \quad (20)$$

$$worst(t) = \max_{j \in \{1, \dots, n\}} fit_j(t) \quad (21)$$

在求解最大值问题时,则定义为:

$$best(t) = \max_{j \in \{1, \dots, n\}} fit_j(t) \quad (22)$$

$$worst(t) = \min_{j \in \{1, \dots, n\}} fit_j(t) \quad (23)$$

将粒子 i 的加速度加到当前的速度上,就可以得到现在的粒子速度和位置:

$$v_i^d(t+1) = rand_i v_i^d(t) + a_i^d(t) \quad (24)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (25)$$

每次迭代后,同步更新粒子的位置和速度,直至达到最大迭代次数或达到要求的精度。

1.5 评价指标

为了更好地反映模型的预测效果。本实验选取均方误差(MSE)和纳什系数(NSE)对测试集的预测结果进行评价。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_0)^2 \quad (26)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - y_0)^2}{\sum_{i=1}^n (y_0 - \bar{y})^2} \quad (27)$$

式中: y_i 为 i 时刻的预测值; y_0 为 i 时刻的实测值; \bar{y} 为实测值的均值; n 为测试集样本数量。

1.6 VMD-GSA-SVR 预测模型

VMD-GSA-SVR 预测模型是一种新型的集成预测模型,其具体的预测流程如图 1 所示。

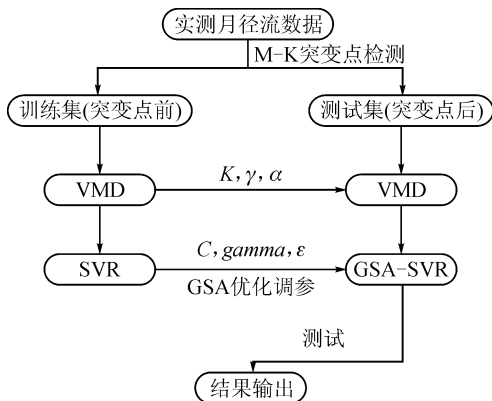


图 1 VMD-GSA-SVR 模型预测流程图

Fig. 1 VMD-GSA-SVR model prediction flow char

1) 对实测月径流数据进行 M-K 突变点检测,突变点前的数据作为训练集,突变点后的数据作为测试集。

2) 对训练集数据进行 VMD 分解,确定模态数 K ,再将测试集数据分解成同样的 K 个模态分量。

3) 将训练集输入 SVR 模型中,运用 GSA 对 SVR 的三个参数 C 、 $gamma$ 、 ϵ 进行全局寻优,用 MSE 作为 GSA 的适应度值公式。

4) 将测试集代入经过参数优化的 GSA-SVR 模型进行预测,并用评价指标进行评价。

2 实例分析

2.1 资料来源与实验环境搭建

咸阳水文站位于陕西省咸阳市东关,始建于 1931 年 6 月 10 日。1957 年 6 月 20 日,其基本水尺断面迁至秦皇路咸阳桥上游一侧(上游 2.6 km 处),更名为咸阳(二)站。该站东经 $108^{\circ}42'$,北纬 $34^{\circ}19'$,控制流域面积 46 827 km²,距河源 606.9 km,距河口 211.1 km。临潼水文站距咸阳站下游 53.7 km,位于临潼区行者乡,东经 $109^{\circ}12'$,北纬 $34^{\circ}26'$,集水面积 97 299 km²。本文选取咸阳站 1956—2010 年 660 个月的实测月径流数据,选取临潼站 1960—2006 年 564 个月的实测月径流数据,如图 2 和图 3 所示。

本实验编程使用 python 3.7,其中 VMD 分解使用第三方库 vmdpy;SVR 使用 sklearn 库;GSA 与对比优化算法使用第三方库 optimal。

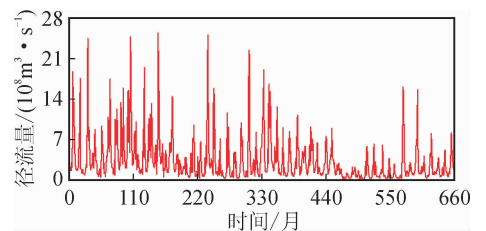


图 2 咸阳站月径流序列

Fig. 2 Monthly runoff series of Xianyang Station

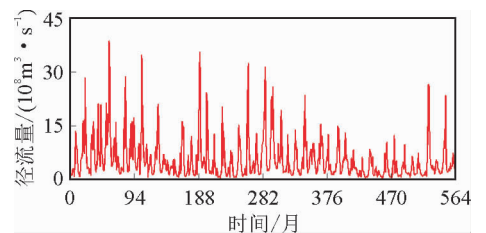


图 3 临潼站月径流序列

Fig. 3 Monthly runoff series of Lintong Station

2.2 训练集与测试集的划分

本实验中,使用 M-K 突变点检测划分训练集和测试集,画出测试统计序列 $\{UF_k\}_{k=2}^n$ 与 $\{UB_k\}_{k=2}^n$,找出两条线在置信区间内的交点,即为突变点,如图 4 和图 5 所示。选择突变点前的实测月径流作为训练集,选择突变点后的实测月径流作为测试集。这样划分可以

更好地测试模型的泛化能力。

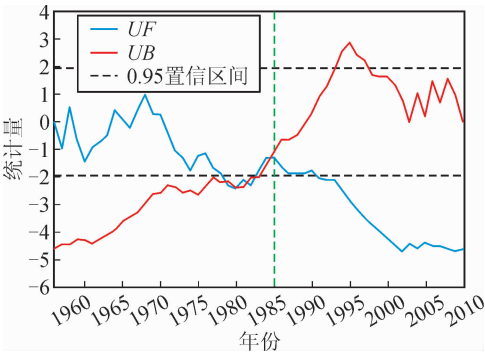


图 4 咸阳站月径流 M-K 突变点检测结果
Fig. 4 Monthly runoff M-K mutation point detection results at Xianyang Station

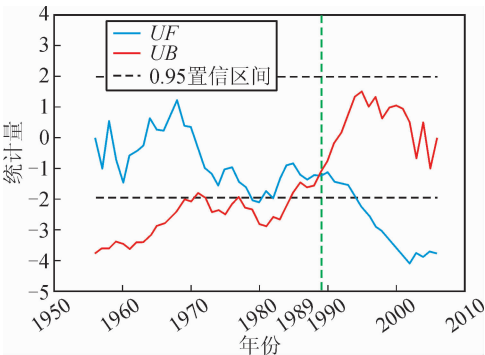


图 5 临潼站月径流 M-K 突变点检测结果
Fig. 5 Monthly runoff M-K mutation point detection results at Lintong Station

由图 4 与图 5 可知,咸阳站的突变点在 1985 年左

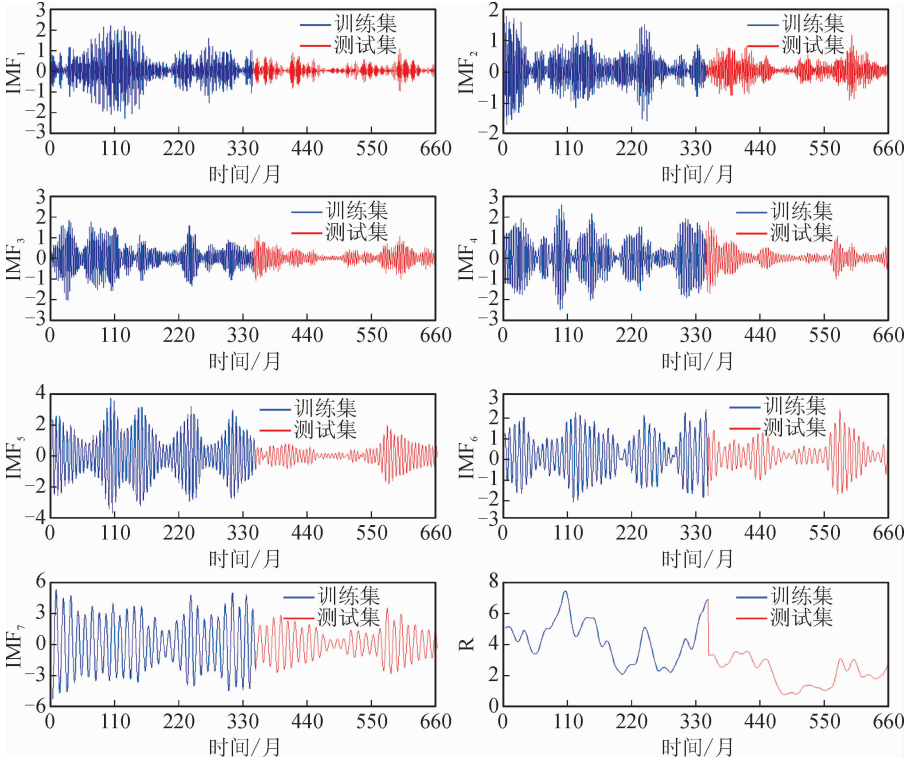


图 6 咸阳站月径流 VMD 分解
Fig. 6 VMD decomposition of monthly runoff at Xianyang Station

右,因此选择 1985 年前 348 个月的实测月径流作为训练集,1985 年后 312 个月的数据作为测试集;临潼站的突变点在 1989 年左右,因此选择 1989 年前 300 个月的实测月径流作为训练集,1989 年后 264 个月的数据作为测试集。

2.3 月径流的 VMD 分解

多数信号分解方法不能手动选择模态数 K ,因此,很多研究在实验开始时就将整个径流序列进行分解,再划分训练集与测试集,这样做实际上将测试集的未知信息带入了模型的训练过程中。鉴于此,可先对训练集进行分解,再选择和训练集相同的模态个数 K ,对测试集进行分解。

确定训练集 VMD 的分解模态数 K 时,采用枚举法,两个站的训练集数据在 $K=8$ 时分解效果最好,且无模态混叠现象^[19]。其余初始参数 γ, α 分别设置为 0、2 000。测试集数据则使用与训练集相同的参数进行分解。

如图 6 与图 7 所示,两个站点的训练集数据和测试集数据分别被分解为 7 个周期性分量(IMF)和 1 个残差(R)。IMF 的频率由大到小排列,R 表示序列的趋势走向。每一个模态都表现出原始序列的特征,使模型能更加准确地学习径流序列的周期性与规律性特征。根据月径流的年际变化规律,将每一个模态分量滞后 12 个月作为模型的输入,原序列的第 13 个月作为输出。在考虑滞后的情况下,咸阳站的训练集样本实际为 336 个月,测试集为 300 个月;临潼站的训练集样本实际为 288 个月,测试集为 252 个月。

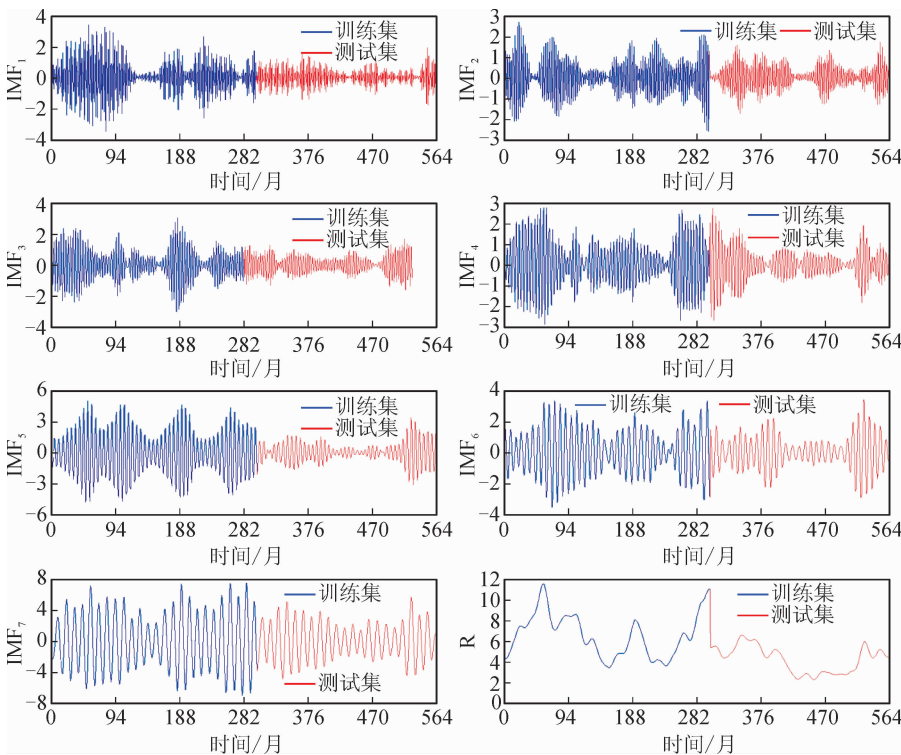


图 7 临潼站月径流 VMD 分解

Fig. 7 VMD decomposition of monthly runoff at Lintong Station

2.4 模型预测及结果评价

为了便于比较,选择无优化调参的 SVR 和 PSO 优化调参的 SVR 作为对比。SVR 的核函数选择 RBF 核函数,其三个参数 C 、 γ 、 ϵ 的搜索区间分别为 $[0.01,100]$ 、 $[0.000\ 001,1]$ 、 $[0.000\ 001,1]$,迭代次数 n 为 100 次,种群规模 $N=40$ 。其中 GSA 的

初始参数 $G_0=100$, α 为 10;PSO 的初始参数 $c_1=1.5$ 、 $c_2=1.7$ 。

输入训练集进行模型训练,得到优化后的 SVR 参数,如表 1 所示。

图 8 和图 9 分别为咸阳站、临潼站测试集预测结果。表 2 为各站预测评价结果。

表 1 优化后的 SVR 参数结果

Tab. 1 Optimized SVR parameter results

站名	SVR 参数	默认值	PSO 优化结果	GSA 优化结果
咸阳站	C	1	80.273 869 1	60.359 521
	γ	0.01	$3.028\ 188\ 73\times10^{-3}$	$1.356\ 711\ 2\times10^{-3}$
	ϵ	0.1	$2.663\ 151\ 69\times10^{-2}$	$2.956\ 224\ 3\times10^{-2}$
临潼站	C	1	34.697 232	73.412 490 1
	γ	0.01	$4.356\ 978\ 5\times10^{-4}$	$6.612\ 359\ 43\times10^{-4}$
	ϵ	0.1	$5.697\ 723\ 1\times10^{-1}$	$6.778\ 257\ 46\times10^{-1}$

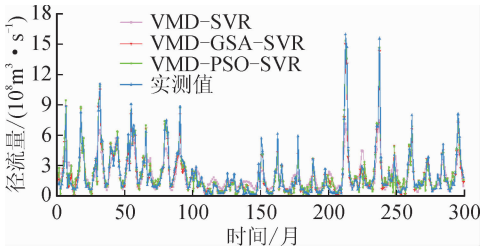


图 8 咸阳站测试集预测结果

Fig. 8 Prediction results of test set at Xianyang Station

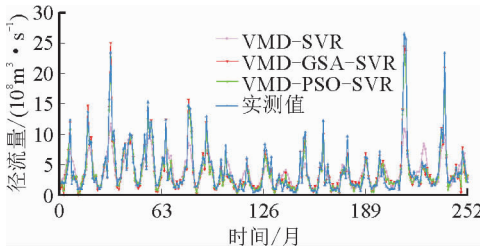


图 9 临潼站测试集预测结果

Fig. 9 Prediction results of test set at Lintong Station

表 2 各站预测评价结果

Tab. 2 Prediction and evaluation results of each station

站名	模型	MSE	NSE
咸阳	VMD-SVR	1. 970 0	0. 654 2
	VMD-PSO-SVR	0. 475 8	0. 930 6
	VMD-GSA-SVR	0. 306 3	0. 946 2
临潼	VMD-SVR	2. 116 5	0. 587 8
	VMD-PSO-SVR	0. 783 8	0. 932 4
	VMD-GSA-SVR	0. 712 1	0. 952 0

利用优化好的参数建立 PSO-SVR、GSA-SVR 模型,将测试集输入模型得出预测结果并进行评价。由图 8 和图 9 可知,在未使用优化算法时,VMD-SVR 模型的拟合效果较 VMD-PSO-SVR 模型、VMD-GSA-SVR 模型明显偏差,这证明了优化算法对 SVR 学习过程的重要性。相较于 VMD-PSO-SVR 模型,VMD-GSA-SVR 模型对峰值与谷值的预测更为精确,表明 GSA 算法对 SVR 参数的寻优能力明显优于 PSO 算法。引力搜索具有良好的全局搜索能力,而粒子群算法易陷入局部最优,因此,通过引力搜索优化的支持向量回归机取得了更优的预测效果。

由表 2 可知,VMD-GSA-SVR 模型对咸阳站月径流预测的 MSE 为 0. 306 3,较 VMD-SVR、VMD-PSO-SVR 模型分别降低了 84%、35%;VMD-GSA-SVR 模型对咸阳站月径流预测的 NSE 为 0. 946 2,较 VMD-SVR、VMD-PSO-SVR 模型分别提高了 45%、2%。VMD-GSA-SVR 模型对临潼站月径流预测的 MSE 为 0. 712 1,较 VMD-SVR、VMD-PSO-SVR 模型分别降低了 66%、9%;VMD-GSA-SVR 模型对临潼站月径流预测的 NSE 为 0. 952 0,较 VMD-SVR、VMD-PSO-SVR 模型分别提高了 61%、2%。由此可知,VMD-GSA-SVR 模型在两个站的预测结果评价中均达到最优,这说明该模型具有良好的拟合精度以及泛化能力。

3 结 论

本文针对月径流小样本数据,建立了 VMD 分解、GSA 与 SVR 组合的 VMD-GSA-SVR 模型,并将其应用于渭河流域咸阳站和临潼站的实测月径流预测中。

1) 利用 M-K 突变点检测划分训练集和测试集,能更好地测试模型对未知数据的拟合程度。

2) 先对训练集进行 VMD 分解,再将训练集的

VMD 参数代入测试集进行分解。这样在分解过程中就不会将测试集的信息带入模型的训练过程,更加符合预测实际。

3) 利用 GSA 对 SVR 的三个参数进行寻优。相较于 PSO-SVR 模型,GSA-SVR 模型在预测中取得了更低的误差以及更高的精度,证明 GSA 是一种可靠有效的优化算法。

4) 将 VMD-GSA-SVR 模型应用于渭河流域咸阳站和临潼站的实测月径流预测中。相较于 VMD-SVR 模型与 VMD-PSO-SVR 模型,VMD-GSA-SVR 模型均取得了最优的预测结果。该模型为渭河流域的月径流预测提供了一条新的途径。

参考文献:

[1] WINDHA M P D, HARYOKO. Optimization of parameter support vector machine (SVM) using genetic algorithm to review Go-Jek’s services[C]// 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2020.

[2] PARK S, BYUN J, LEE J, et al. HE-friendly algorithm for privacy-preserving SVM training[J]. IEEE Access, 2020, 8: 57414-57425.

[3] 黄巧玲, 粟晓玲. 基于小波支持向量机耦合的月径流预测方法[J]. 水力发电学报, 2015, 34(3): 1-7, 20. HUANG Qiaoling, SU Xiaoling. Wavelet support vector machine-coupling method for monthly runoff forecasting [J]. Journal of Hydroelectric Engineering, 2015, 34(3): 1-7, 20.

[4] 周有荣, 崔东文. 基于混合核 SHTS-SVM 的年径流预测[J]. 水资源与水工程学报, 2019, 30(3): 66-72. ZHOU Yourong, CUI Dongwen. Annual runoff prediction based on mixed kernel simultaneous heat transfer search-support vector machine [J]. Journal of Water Resources and Water Engineering, 2019, 30 (3): 66-72.

[5] 李祥蓉. 静电放电算法-混合核 SVM 的月径流预报模

- 型及应用[J]. 人民珠江, 2020, 41(1): 23-28.
- LI Xiangrong. Monthly runoff forecasting model based on electrostatic discharge algorithm-mixed kernel SVM and its application [J]. Pearl River, 2020, 41(1): 23-28.
- [6] 王迁, 杨明祥, 雷晓辉, 等. 基于 PSO-SVR 的丹江口年径流预报[J]. 南水北调与水利科技, 2018, 16(3): 65-71.
- WANG Qian, YANG Mingxiang, LEI Xiaohui, et al. Annual runoff forecast for Danjiangkou based on PSO-SVR [J]. South-to-North Water Transfers and Water Science & Technology, 2018, 16(3): 65-71.
- [7] 梁浩, 黄生志, 孟二浩, 等. 基于多种混合模型的径流预测研究[J]. 水利学报, 2020, 51(1): 112-125.
- LIANG Hao, HUANG Shengzhi, MENG Erhao, et al. Runoff prediction based on multiple hybrid models [J]. Journal of Hydraulic Engineering, 2020, 51(1): 112-125.
- [8] 王瑞平, 陈杰, 山世光, 等. 基于支持向量机的人脸检测训练集增强[J]. 软件学报, 2008, 19(11): 2921-2931.
- WANG Ruiping, CHEN Jie, SHAN Shiguang, et al. Enhancing training set for face detection based on SVM [J]. Journal of Software, 2008, 19(11): 2921-2931.
- [9] 刘磊, 杨晔, 刘赛, 等. 基于生存理论训练机器学习的智能驾驶路径生成方法[J]. 控制与决策, 2020, 35(10): 2433-2441.
- LIU Lei, YANG Ye, LIU Sai, et al. Intelligent driving path generation method for intelligent driving based on machine learning trained by viability theory [J]. Control and Decision, 2020, 35(10): 2433-2441.
- [10] 史贵君, 郭力源, 林涛, 等. 深圳市降水变化特征及其未来变化情况预测[J]. 长江科学院院报, 2020, 37(5): 28-33.
- SHI Guijun, GUO Liyuan, LIN Tao, et al. Variation characteristics and future changes of precipitation in Shenzhen [J]. Journal of Yangtze River Scientific Research Institute, 2020, 37(5): 28-33.
- [11] 金丰, 邵清. 信号分解与融合神经网络的金融数据预测研究[J]. 小型微型计算机系统, 2020, 41(6): 1140-1146.
- JIN Feng, SHAO Qing. Research on financial data forecast based on signal decomposition and fusion neural network [J]. Journal of Chinese Computer Systems, 2020, 41(6): 1140-1146.
- [12] ZHANG Y, GAO S, PAN L, et al. Fusion of GNSS and speedometer based on VMD and its application in bridge deformation monitoring[J]. Sensors, 2020, 20(3): 694.
- [13] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [14] 吴晋武, 张海峰, 冉旭东. 基于数据约减和支持向量机的非参数回归短时交通流预测算法[J]. 公路交通科技, 2020, 37(7): 129-134.
- WU Jinwu, ZHANG Haifeng, RAN Xudong. Non-parametric regressive short-term traffic flow forecast algorithm based on data reduction and SVM [J]. Journal of Highway and Transportation Research and Development, 2020, 37(7): 129-134.
- [15] 张楠, 夏自强, 江红. 基于多因子量化指标的支持向量机径流预测[J]. 水利学报, 2010, 41(11): 1318-1324.
- ZHANG Nan, XIA Ziqiang, JIANG Hong. Prediction of runoff based on the multiple quantity index of SVM [J]. Journal of Hydraulic Engineering, 2010, 41(11): 1318-1324.
- [16] ZEMKHO A B, ZHOU S. Theoretical and numerical comparison of the Karush-Kuhn-Tucker and value function reformulations in bilevel optimization [J]. Computational Optimization and Applications, 2021, 78: 625-674.
- [17] 代兴兰. 回归支持向量机集成模型在年径流预测中的应用[J]. 长江科学院院报, 2015, 32(4): 12-17.
- DAI Xinglan. Application of SVR ensemble model to annual runoff forecasting [J]. Journal of Yangtze River Scientific Research Institute, 2015, 32(4): 12-17.
- [18] 许瀚誉, 冯翔, 虞慧群. 万有引力与群体状态自适应的智能优化算法[J]. 计算机工程与应用, 2020, 56(9): 48-55.
- XU Hanyu, FENG Xiang, YU Huiqun. Novel optimization algorithm based on universal gravitational and colon state adaptation [J]. Computer Engineering and Applications, 2020, 56(9): 48-55.
- [19] 陈永高, 钟振宇. 基于改进 EEMD 算法的桥梁结构响应信号模态分解研究[J]. 振动与冲击, 2019, 38(10): 23-30.
- CHEN Yonggao, ZHONG Zhenyu. Modal decomposition of response signals for a bridge structure based on the improved EEMD [J]. Journal of Vibration and Shock, 2019, 38(10): 23-30.